UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

CAMBRIDGE WORKING PAPERS IN ECONOMICS
CAMBRIDGE-INET WORKING PAPERS

Estimation of Common Factors for Microstructure Noise and Efficient Price in a High-frequency Dual Factor Model

Yu-Ning Li
University of York

Jia Chen
University of York

Oliver Linton
University of Cambridge

Abstract

We develop the Double Principal Component Analysis (DPCA) based on a dual factor structure for high-frequency intraday returns data contaminated with microstructure noise. The dual factor structure allows a factor structure for the microstructure noise in addition to the factor structure for efficient log-prices. We construct estimators of factors for both efficient log-prices and microstructure noise as well as their common components, and provide uniform consistency of these estimators when the number of assets and the sampling frequency go to infinity. In a Monte Carlo exercise, we compare our DPCA method to a PCA-VECM method. Finally, an empirical analysis of intraday returns of S&P 500 Index constituents provides evidence of co-movement of the microstructure noise that distinguishes from latent systematic risk factors.

Reference Details

2150	Cambridge Working Papers in Economics
2021/22	Cambridge-INET Working Paper Series
Published	30 June 2021
Key Words	Cointegration, Factor model, High-frequency data, Microstructure noise, Non-stationarity
JEL Codes	C10, C13, C14, C33, C38
Websites	www.econ.cam.ac.uk/cwpe www.inet.econ.cam.ac.uk/working-papers

Estimation of Common Factors for Microstructure Noise and Efficient Price in a High-frequency Dual Factor Model

Yu-Ning Li[a], Jia Chen[a], Oliver Linton[b]

[a] *Department of Economics and Related Studies, University of York, UK*
[b] *Faculty of Economics, University of Cambridge, UK*

Abstract

We develop the Double Principal Component Analysis (DPCA) based on a dual factor structure for high-frequency intraday returns data contaminated with microstructure noise. The dual factor structure allows a factor structure for the microstructure noise in addition to the factor structure for efficient log-prices. We construct estimators of factors for both efficient log-prices and microstructure noise as well as their common components, and provide uniform consistency of these estimators when the number of assets and the sampling frequency go to infinity. In a Monte Carlo exercise, we compare our DPCA method to a PCA-VECM method. Finally, an empirical analysis of intraday returns of S&P 500 Index constituents provides evidence of co-movement of the microstructure noise that distinguishes from latent systematic risk factors.

Keywords: Cointegration, Factor model, High-frequency data, Microstructure noise, Non-stationarity, *AMS Subject Classification*:

Preprint submitted to Journal of *June 30, 2021*

1. Introduction

Factor models are widely used in financial econometrics. Their popularity is partly due to the easiness of their implementation and their effectiveness in dimension reduction. More and more observable factors have been investigated and reported (see, e.g., Ross (1976), Sharpe (1994), Fama and French (1993, 2015) and Carhart (1997)) as driving stock returns. Researchers have also found common components in other attributes of financial assets such as volatility and liquidity. For example, Chordia et al. (2000) document the commonality in liquidity, which remains significant after controlling for volatility, volume, and price. The factor structure is not found in isolation. Indeed, price and other attributes of stocks have been found to have correlated common factors. Hasbrouck and Seppi (2001) use principal component analysis to show that common factors exist in order flows and equity returns. In addition, using canonical correlation analysis, they find that the common factor in returns is highly correlated with the common factor in order flows. Hallin and Liška (2011) propose a two-step general dynamic factor method to account for a joint factor structure of sub-panels, which is further developed by Barigozzi and Hallin (2016) and Barigozzi and Hallin (2017) for extracting the market volatility shocks. They find that returns and volatilities can be decomposed into four mutually orthogonal components: a strongly idiosyncratic component, a strongly common component, a weakly common component, and a weakly idiosyncratic component.

The increasing availability of high-frequency transaction data motivates applying this methodology to intraday stock prices. However, this setting raises certain theoretical and computational challenges. Compared to the discrete-time factor model, new mathematical tools are required to deal with a continuous-time setting, where long-span asymptotics (also called increasing domain asymptotics) gives way to infill asymptotics (also called fixed domain asymptotics). For example, Fan et al. (2016) and Aït-Sahalia and Xiu (2017) extend Fan et al. (2013)'s Principal Orthogonal complEment Thresholding (POET) method to high-frequency factor models. Market microstructure is an additional challenge that must be faced. The specifics of market organization and market participants' behaviour induce certain short-run patterns in security prices. These patterns, such as bid-ask bounce and price-discreteness, lead to a deviation from the fundamental values (also known as efficient prices) of the securities. The security prices are thereby contaminated with market microstructure noise, which affects the estimation of parameters of interest such as volatility. Market microstructure models have been used to capture a variety of frictions inherent in the trading the efficient price process. Roll (1984) is among the first to propose a dichotomous structure in which the observed market price is the sum of the efficient price and an exogenous i.i.d. bid-ask spread. After that, Hasbrouck and Ho (1987), Choi et al. (1988) and Hasbrouck (1993) consider extended models with positive dependence in bid and ask transactions. More complicated price patterns arising from microstructure noise, such as asynchronous trading, have been investigated by researchers under the fundamental dichotomous structure.

High-dimensional models with microstructure noise have been developed more recently. Wang and Zou (2010) propose the first noise-robust estimators for the integrated volatility matrix and establish an asymptotic theory that allows both the sample size and the number of assets to approach infinity, see also Tao et al. (2011, 2013a,b), and Kim et al. (2016) for related results. However, these papers assume that the integrated volatility matrix is sparse, which often contradicts our intuition from low frequency data analysis. To solve this problem, Pelger (2019) and Dai et al. (2019) develop a continuous-time factor model with microstructure noise. Likewise, Bollerslev et al. (2019) investigate a continuous-time factor model and assume that the microstructure noise can have a factor structure itself. They use the modulated realized covariance estimator (henceforth MRC) of Christensen et al. (2010) to eliminate the effect of the microstructure noise on the estimation without explicitly estimating the factors of the microstructure noise and separating them from those of the efficient prices. They establish the consistency and bound the rate of convergence of the estimated integrated covariance matrix of the efficient price process in the large dimensional case. Related to this, Pelger (2019) classifies factors in a high-frequency factor model into jump factors and continuous factors.

We consider the dual factor model of Bollerslev et al. (2019) but we take a different approach to estimation. Our goal is to identify and separate the factors and common components from both sources: the efficient price process and the microstructure noise process. Factors for the efficient prices arise from information about future security cash flows and thereby are long-lasting, whereas factors for the microstructure noise are transient and due to the nature of trading behaviour; both are of interest. We develop a methodology that is inspired by Bai and Ng (2004), who propose a test procedure called Panel Analysis of Non-stationarity in Idiosyncratic and Common Components (PANIC), which can be used to identify non-stationary factors in discrete time series. We extend the PANIC approach to our high-frequency dual factor model where the concept of co-integration cannot be used. Our methodology is in three parts. First, we estimate the common factors and loadings of both signal and noise components simultaneously from the observed return matrix. The PCA method identifies factors and idiosyncratic errors by the eigenvalues. The common factors have eigenvalues that diverge at a rate of d, where d is the number of assets, and the idiosyncratic errors have bounded eigenvalues. There is a big gap from $O(d)$ to $O(1)$ so that even if the efficient returns are of a smaller order, we are still able to identify their (weak) factors in the large dimensional case. The second step separates the return factors into the efficient price factors and the microstructure noise factors. This involves a second PCA step on the cumulated factors found in the first step, following the approach of Bai and Ng (2004). The final step is to cumulate the return factors to define the common factors in prices. We establish the consistency of our procedures as the number of assets increases and the number of infill observations for each asset increases. Our asymptotic framework allows for a rich diversity in the relative size of the efficient price process and the microstructure noise and in the relative size of the common component of the microstructure noise and the idiosyncratic components of the noise. This is important because a number of authors have documented that in frequently traded assets, the microstructure noise component can be quite small. Also, the Epps effect, whereby observed cross asset correlations shrink with sampling frequency, can be captured in our framework when the idiosyncratic component of the noise is larger element by element than the common component. Our model allows the so-called "weak factors", c.f., Briggs and MacCallum (2003) and Onatski (2010). We provide a full analysis of the convergence rates of all our estimators, which are affected by the magnitudes of the components of microstructure noise. To determine the number of factors we use several methods proposed in the discrete time literature and investigate their performance on simulated data. We find that the Bai and Ng (2002) information criterion performs well in terms of selecting the total number of factors and the PANIC test works well in our setting for determining the number of factors in the efficient price process. We apply our method to the intraday returns of S&P 500 Index constituents. The empirical analysis provides evidence of co-movement of the microstructure noise.

The rest of the paper is organized as follows. Section 2 specifies the model and its assumptions. Section 3 proposes the high-frequency PANIC estimation procedure and presents the asymptotic properties for the estimators. Section 3.3 provides finite-sample simulation results and Section 5 demonstrates the applicability of our proposed method through an empirical study. Section 6 concludes. The proofs of our main results are relegated to the Appendix.

Throughout the paper, we use $||\cdot||_2$ to denote the Euclidean norm of a vector. For a real symmetric matrix $\mathbf{S}$, we denote its k-th largest eigenvalue and trace by $\mu_k(\mathbf{S})$ and $\mathsf{tr}(\mathbf{S})$, respectively. For any $m \times n$ matrix $\mathbf{M} = (m_{ij})$, let $||\mathbf{M}||_{\mathsf{sp}}$, $||\mathbf{M}||_1$, $||\mathbf{M}||_\infty$, $||\mathbf{M}||_F$ and $||\mathbf{M}||_{\mathsf{MAX}}$ denote the spectral norm, the l_1 norm, the l_∞ norm, the Frobenius norm, and the max norm of $\mathbf{M}$, respectively. Specifically, $||\mathbf{M}||_{\mathsf{sp}} = \sqrt{\mu_1(\mathbf{M}^\intercal\mathbf{M})}$, $||\mathbf{M}||_1 = \max_j \sum_i |m_{ij}|$, $||\mathbf{M}||_\infty = \max_i \sum_j |m_{ij}|$, $\|\mathbf{M}\|_F = \sqrt{\mathsf{tr}(\mathbf{M}^\intercal\mathbf{M})} = \sqrt{\sum_{i,j} m_{ij}^2}$ and $||\mathbf{M}||_{\mathsf{MAX}} = \max_{i,j} |m_{ij}|$. Let $\mathbf{1}_n$ denote an n-dimensional vectors of 1's and $\mathbf{L}_n$ an n-by-n lower triangular matrix where all diagonal and below-diagonal entries are 1's. Also let $a \vee b$ and $a \wedge b$ denote $\max\{a, b\}$ and $\min\{a, b\}$, and x_+ and x_- denote $\max\{0, x\}$ and $\min\{0, x\}$, respectively.

2. Model Setup and Assumptions

2.1. Dual factor structure

Let X_{it} denote the observed log transaction price of stock i at time t, for $i = 1, ..., d$. We allow d to diverge with n, although we have suppressed the subscript n for d. For the sake of simplicity, we assume that price observations of all stocks are synchronously collected, and that price observations for each stock are equidistantly collected in the fixed time interval $[0, T]$. Thus we do not consider non-synchronous trading explicitly. Without loss of generality, we let $T = 1$. Let n be the number of observations and $\Delta = 1/n$. Then, the prices are observed at the time points $t = 0, \Delta, 2\Delta, \ldots, n\Delta$. Although we assume the length of the time interval between any two successive transactions to be equal, our results still hold under the more general assumption that the length of the time interval is unequal but is uniformly of order $1/n$.

We assume that the observed log transaction price, X_{it}, can be decomposed into the unobserved efficient log-price X_{it}^* plus a noise component Z_{it}, i.e.,

$$X_{it} = X_{it}^* + Z_{it} \quad \text{or} \quad \boldsymbol{X}_t = \boldsymbol{X}_t^* + \boldsymbol{Z}_t, \tag{2.1}$$

where $\boldsymbol{X}_t^* = (X_{1t}^*, \ldots, X_{dt}^*)^\intercal$ and $\boldsymbol{Z}_t = (Z_{1t}, \ldots, Z_{dt})^\intercal$. For each component of X_{it}, we introduce a factor structure (see Assumptions 1 and 2 below) and therefore, name the model as a *dual factor model*.

***Assumption* 1.** *(Factor Structure for Efficient Log-price)*
(i) The efficient log-price $\boldsymbol{X}_t^*$ follows a factor model of the form,

$$\begin{cases} \mathsf{d}\boldsymbol{X}_t^* &= \boldsymbol{\Lambda}_F \mathsf{d}\boldsymbol{F}_t + \mathsf{d}\boldsymbol{U}_t, \\ \mathsf{d}\boldsymbol{F}_t &= \boldsymbol{\sigma}_{Ft}\mathsf{d}\boldsymbol{B}_t^F, \\ \mathsf{d}\boldsymbol{U}_t &= \boldsymbol{\sigma}_{Ut}\mathsf{d}\boldsymbol{B}_t^U, \end{cases}$$

where $\boldsymbol{\Lambda}_F = (\lambda_{F,ik})_{1\le i\le d, 1\le k\le K_F}$ denotes the $d \times K_F$ matrix of factor loadings, K_F is the number of factors, $\boldsymbol{F}_t = (F_{1t}, ..., F_{K_F t})^\intercal$ denotes latent factors, $\boldsymbol{U}_t = (U_{1t}, ..., U_{dt})^\intercal$ is the idiosyncratic component, $\boldsymbol{\sigma}_{Ft}$ is a $K_F \times K_F$ càd-làg spot volatility matrix for factors, $\boldsymbol{\sigma}_{Ut}$ is a $d \times d$ càd-làg spot volatility matrix for idiosyncratic errors, and $\boldsymbol{B}_t^F = (B_{1t}^F, \ldots, B_{K_F t}^F)^\intercal$ and $\boldsymbol{B}_t^U = (B_{1t}^U, \ldots, B_{dt}^U)^\intercal$ are independent Brownian motions.
(ii) The initial states of $\boldsymbol{F}_t$ and $\boldsymbol{U}_t$ satisfy $\|\boldsymbol{F}_0\|_{\mathsf{MAX}} = O_P(1)$ and $\|\boldsymbol{U}_0\|_{\mathsf{MAX}} = O_P(1)$.
(iii) Denote the spot covariance matrices of $\boldsymbol{F}_t$ and $\boldsymbol{U}_t$ by $\boldsymbol{\Sigma}_{Ft} = \boldsymbol{\sigma}_{Ft}\boldsymbol{\sigma}_{Ft}^\intercal$ and $\boldsymbol{\Sigma}_{Ut} = \boldsymbol{\sigma}_{Ut}\boldsymbol{\sigma}_{Ut}^\intercal$, and define $\boldsymbol{\Sigma}_{Ft-} = \lim_{\Delta\to 0}\boldsymbol{\Sigma}_{F,t-\Delta}$ and $\boldsymbol{\Sigma}_{Ut-} = \lim_{\Delta\to 0}\boldsymbol{\Sigma}_{U,t-\Delta}$. Then, $\boldsymbol{\Sigma}_{Ft}$, $\boldsymbol{\Sigma}_{Ft-}$, $\boldsymbol{\Sigma}_{Ut}$ and $\boldsymbol{\Sigma}_{Ut-}$ are all positive-definite and satisfy

$$\max\{\|\boldsymbol{\Sigma}_{Ft}\|_{\mathsf{MAX}},\ \|\boldsymbol{\Sigma}_{Ft-}\|_{\mathsf{MAX}},\ \|\boldsymbol{\Sigma}_{Ut}\|_{\mathsf{MAX}},\ \|\boldsymbol{\Sigma}_{Ut-}\|_{\mathsf{MAX}}\} \le C_\sigma$$

for all $t \in [0, 1]$, where C_σ is a positive constant independent of n and d.

***Remark* 2.1.** *Our dual factor model follows the setting of Bollerslev et al. (2019), and inherits several limitations. Firstly, we do not allow a drift term in the diffusion model. More general settings can be seen in Dai et al. (2019) and Barigozzi et al. (2020b) for instance. Secondly, we assume that the factor loadings are constant and thus, neither random nor time-varying loadings are allowed. We refer to Aït-Sahalia et al. (2020) for high-frequency factor models with time-varying betas. Thirdly, we assume the number of factors is fixed. We refer to Fan et al. (2011, 2016) for factor models with the number of factors increasing with d. Another limitation is that jumps are excluded in our model. Dealing with jumps would require more complicated procedures and hence, we leave it for future work.*

Assumption 2. *(Factor Structure for Market Microstructure Noise)*
The microstructure noise $\boldsymbol{Z}_t$ follows a factor model whose magnitude may depend on the sampling frequency, that is

$$\boldsymbol{Z}_t = \boldsymbol{\Lambda}_G \mathbf{D}_G \boldsymbol{G}_t + \mathbf{D}_V \boldsymbol{V}_t, \tag{2.2}$$

where $\boldsymbol{\Lambda}_G = (\lambda_{G,ik})_{1\le i\le d, 1\le k\le K_G}$ denotes the $d \times K_G$ matrix of factor loadings with K_G being the number of factors for microstructure noise, $\boldsymbol{G}_t = (G_{1t}, ..., G_{K_G t})^\intercal$ denotes the latent factors, $\boldsymbol{V}_t = (V_{1t}, ..., V_{dt})^\intercal$ is the vector of idiosyncratic components, and $\mathbf{D}_G$ and $\mathbf{D}_V$ are two diagonal matrices satisfying $\mu_1(\mathbf{D}_G) = O(n^{\bar{\tau}_G})$, $\mu_1(\mathbf{D}_G^{-1}) = O(n^{-\underline{\tau}_G})$ and $\mu_1(\mathbf{D}_V) = O(n^{\bar{\tau}_V})$, where $\bar{\tau}_G$, $\underline{\tau}_G$, and $\bar{\tau}_V$ are some constants.

The introduction of $\mathbf{D}_G$ and $\mathbf{D}_V$ in (2.2) allows the microstructure noise to be larger or smaller in magnitude than the efficient log-prices (depending on the values of $\bar{\tau}_G$ and $\bar{\tau}_V$). With such a setup, the magnitude of $\boldsymbol{\Lambda}_G$, $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$ is independent of n. A similar treatment can be found in Kim et al. (2016). Our model is an extension of the model of Bollerslev et al. (2019), who only consider $\mathbf{D}_G = \mathbf{I}_{K_G}$ and $\mathbf{D}_V = \mathbf{I}_d$. However, they use a similar setting when generating simulation data (also see Section 3.3) without discussing the asymptotic impacts of $\mathbf{D}_G$ and $\mathbf{D}_V$.

To introduce the first-differenced form of the dual factor model, we use little letters to denote the first-order differences of random variables. Specifically, define the return as $\boldsymbol{x}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-\Delta}$, and the efficient return (or frictionless return) as $\boldsymbol{x}_t^* = \int_{t-\Delta}^t \mathsf{d}\boldsymbol{X}_s^* = \boldsymbol{X}_t^* - \boldsymbol{X}_{t-\Delta}^*$. Denote $\boldsymbol{f}_t = \int_{t-\Delta}^t \boldsymbol{\sigma}_{fs} \mathsf{d}\boldsymbol{B}_s^F = \boldsymbol{F}_t - \boldsymbol{F}_{t-\Delta}$, $\boldsymbol{g}_t = \boldsymbol{G}_t - \boldsymbol{G}_{t-\Delta}$, $\boldsymbol{z}_t = \boldsymbol{Z}_t - \boldsymbol{Z}_{t-\Delta}$, $\boldsymbol{u}_t = \int_{t-\Delta}^t \boldsymbol{\sigma}_{Us} \mathsf{d}\boldsymbol{B}_s^U = \boldsymbol{U}_t - \boldsymbol{U}_{t-\Delta}$, and $\boldsymbol{v}_t = \boldsymbol{V}_t - \boldsymbol{V}_{t-\Delta}$. Then by (2.1)–(2.2), we can write the dual factor model as

$$\begin{cases} \boldsymbol{x}_t &= \boldsymbol{x}_t^* + \boldsymbol{z}_t \\ \boldsymbol{x}_t^* &= \boldsymbol{\Lambda}_F \boldsymbol{f}_t + \boldsymbol{u}_t \\ \boldsymbol{z}_t &= \boldsymbol{\Lambda}_G \mathbf{D}_G \boldsymbol{g}_t + \mathbf{D}_V \boldsymbol{v}_t \end{cases} \tag{2.3}$$

Combining the factor structures for $\boldsymbol{x}_t^*$ and $\boldsymbol{z}_t$, we have

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{\Lambda}_F \boldsymbol{f}_t + \boldsymbol{u}_t + \boldsymbol{\Lambda}_G \mathbf{D}_G \boldsymbol{g}_t + \mathbf{D}_V \boldsymbol{v}_t, \\ &= \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{h}_t + \boldsymbol{w}_t, \end{aligned} \tag{2.4}$$

where $\boldsymbol{h}_t = (\boldsymbol{f}_t^\intercal, n^{-1/2}\boldsymbol{g}_t^\intercal)^\intercal$, $\boldsymbol{\Lambda}_H = (\boldsymbol{\Lambda}_F, \boldsymbol{\Lambda}_G)$, $\boldsymbol{w}_t = \boldsymbol{u}_t + \mathbf{D}_V \boldsymbol{v}_t$ and $\mathbf{D}_H = \mathsf{diag}(\mathbf{I}_{K_F}, n^{1/2}\mathbf{D}_G)$. This can be seen as a factor structure for $\boldsymbol{x}_t$ with $\boldsymbol{h}_t = (\boldsymbol{f}_t^\intercal, n^{-1/2}\boldsymbol{g}_t^\intercal)^\intercal$ being the factors and $\boldsymbol{\Lambda}_H = (\boldsymbol{\Lambda}_F, \boldsymbol{\Lambda}_G)$ being the factor loadings. Note that $\boldsymbol{g}_t$ is divided by $n^{1/2}$ in $\boldsymbol{h}_t$ so that both components of $\boldsymbol{h}_t$ are of the same magnitude. Consequently, the magnitude matrix $\mathbf{D}_G$ is multiplied by $n^{1/2}$ in $\mathbf{D}_H$. This also gives more insight into the role $\mathbf{D}_G$ plays. If $\bar{\tau}_G > -1/2$, the factors of the microstructure noise dominate those of the efficient returns and vice versa. Similarly, since $\boldsymbol{u}_t$ and $n^{-1/2}\boldsymbol{v}_t$ are of the same magnitude, if $\bar{\tau}_V > -1/2$, the idiosyncratic components of the microstructure noise dominates those of the efficient returns and vice versa.

Denote the number of independent factors in factor model (2.4) as K_H. If $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$ are collinear, K_H will be less than $K_F + K_G$. In such a case, the efficient prices and microstructure noise are not separable and the dichotomous structure fails. Thus for identification purposes, we exclude this situation and make the following assumption .

***Assumption* 3.** *(Independence between efficient prices and microstructure noise) The discrete time series $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$ are independent of the continuous-time processes $\boldsymbol{F}_t$ and $\boldsymbol{U}_t$.*

***Assumption* 4.** *(Factor loadings) The factor loadings matrix $\boldsymbol{\Lambda}_H$ satisfies*

$$\|\boldsymbol{\Lambda}_H\|_{\mathsf{MAX}} < C_\Lambda, \qquad \|\boldsymbol{\Lambda}_H^\intercal \boldsymbol{\Lambda}_H / d\|_{\mathsf{sp}} < C_\Lambda, \qquad \|(\boldsymbol{\Lambda}_H^\intercal \boldsymbol{\Lambda}_H / d)^{-1}\|_{\mathsf{sp}} < C_\Lambda,$$

where C_Λ is a positive constant independent of n and d.

***Remark* 2.2.** *If $\boldsymbol{\sigma}_{Ft}$ is a constant spot volatility matrix which does not vary with t, then we can use the notion of cointegration as in Bai and Ng (2004). In this case, the cointegration rank of $\boldsymbol{H}_t := (\boldsymbol{F}_t^\intercal, n^{-1/2}\boldsymbol{G}_t^\intercal)^\intercal$ is $K_H - K_F = K_G$, which is the number of factors for microstructure noise.*

***Remark* 2.3.** *It is prevalent to assume independence between price components due to fundamental security value and noise attributable to market rules and trading mechanisms in a univariate microstructure model. One of the reasons is modelling simplicity, so that the two components of the model can be identified and the model can have sensible economic and statistical representation and interpretation. However, the independence assumption may sometimes be violated. For example, it is shown in Delattre and Jacod (1997) and Li and Mykland (2007) that this assumption can be substantially weakened, especially for processes contaminated with rounding errors. Glosten (1987) and Glosten and Harris (1988) also show that the microstructure noise may no longer be uncorrelated with the efficient price when asymmetric information is involved. Recently, there has been some research that allows correlation between them, such as Kunitomo and Kurisu (2019), who modify the Separating Information Maximum Likelihood due to Kunitomo and Sato (2013) to detect hidden factors of quadratic variation in the presence of correlated market microstructure noise.*

For easy reference, we summarise the notation used for variables and factors in Table 1. We use different fonts to distinguish between matrices, vectors and scalars. For example, $\mathscr{X} = (\boldsymbol{X}_\Delta, \ldots, \boldsymbol{X}_{n\Delta})^\intercal$ is an $n \times d$ matrix of observed prices, $\boldsymbol{X}_{s\Delta}^\intercal$ is its s-th row, and $X_{i,s\Delta}$ is the (s,i)-entry of the matrix $\mathscr{X}$. Following the same rule, other variables are defined analogously.

2.2. Covariance structure

Define the integrated covolatility matrix of $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ as $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_U$, respectively. That is $\boldsymbol{\Sigma}_F = \int_0^1 \boldsymbol{\Sigma}_{Ft} \mathsf{d}t$ and $\boldsymbol{\Sigma}_U = \int_0^1 \boldsymbol{\Sigma}_{Ut} \mathsf{d}t$. Then the factor structure for efficient prices leads to the following identity,

$$\boldsymbol{\Sigma}_{x^*} = \boldsymbol{\Lambda}_F \boldsymbol{\Sigma}_F \boldsymbol{\Lambda}_F^\intercal + \boldsymbol{\Sigma}_U, \tag{2.5}$$

Table 1: Notation for variables/factors in the dual factor model

Variables	Aggregation Form			First-difference Form		
	Matrix -wise	Row -wise	Element -wise	Matrix -wise	Row -wise	Element -wise
Observed price	$\mathscr{X}$	$\boldsymbol{X}_t$	X_{it}	$\mathscr{x}$	$\boldsymbol{x}_t$	x_{it}
Efficient price (EP)	$\mathscr{X}^*$	$\boldsymbol{X}_t^*$	X_{it}^*	$\mathscr{x}^*$	$\boldsymbol{x}_t^*$	x_{it}^*
Microstructure noise (MN)	$\mathscr{Z}$	$\boldsymbol{Z}_t$	Z_{it}	$\mathscr{z}$	$\boldsymbol{z}_t$	z_{it}
Factors for EP	$\mathscr{F}$	$\boldsymbol{F}_t$	F_{jt}	$\mathscr{f}$	$\boldsymbol{f}_t$	f_{jt}
Factors for MN	$\mathscr{G}$	$\boldsymbol{G}_t$	G_{jt}	$\mathscr{g}$	$\boldsymbol{g}_t$	g_{jt}
Total factors	$\mathscr{H}$	$\boldsymbol{H}_t$	H_{jt}	$\mathscr{h}$	$\boldsymbol{h}_t$	h_{jt}
Idiosyncratic errors for EP	$\mathscr{U}$	$\boldsymbol{U}_t$	U_{it}	$\mathscr{u}$	$\boldsymbol{u}_t$	u_{it}
Idiosyncratic errors for MN	$\mathscr{V}$	$\boldsymbol{V}_t$	V_{it}	$\mathscr{v}$	$\boldsymbol{v}_t$	v_{it}
Total idiosyncratic errors	$\mathscr{W}$	$\boldsymbol{W}_t$	W_{it}	$\mathscr{w}$	$\boldsymbol{w}_t$	w_{it}

[1] The first dimension of the matrices in the table are set as time, while the second dimension are set as a stock or a factor. We have $t = \Delta, \cdots, n\Delta$, $1 \le i \le d$ and $1 \le j \le K$, where $K = K_F$, K_G, or K_H, depending on the circumstance. Note that in the subscript of the element-wise notation, we write the column index first.

[2] Although when $t = 0$, $\boldsymbol{X}_t$ is observable, t starts from Δ in the matrix-wise notation for both aggregation form and first-difference form, for the sake of consistency.

where $\boldsymbol{\Sigma}_{x^*}$ is the integrated covolatility matrix of $\boldsymbol{x}_t^*$. Similarly, the factor structure for microstructure noise leads to the following identity,

$$\boldsymbol{\Sigma}_z = \boldsymbol{\Lambda}_G \mathbf{D}_G \boldsymbol{\Sigma}_g \mathbf{D}_G \boldsymbol{\Lambda}_G^\intercal + \mathbf{D}_V \boldsymbol{\Sigma}_v \mathbf{D}_V. \tag{2.6}$$

where $\boldsymbol{\Sigma}_z = \mathsf{Var}(\boldsymbol{z}_t)$, $\boldsymbol{\Sigma}_g = \mathsf{Var}(\boldsymbol{g}_t)$, and $\boldsymbol{\Sigma}_v = \mathsf{Var}(\boldsymbol{v}_t)$.[1] For the integrated covolatility matrix of observed prices, we combine (2.5) and (2.6) to obtain

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\Sigma}_h \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal + \boldsymbol{\Sigma}_w, \tag{2.7}$$

where $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_{x^*} + n\boldsymbol{\Sigma}_z$, $\boldsymbol{\Sigma}_h = \mathsf{diag}(\boldsymbol{\Sigma}_F, \boldsymbol{\Sigma}_g)$, and $\boldsymbol{\Sigma}_w = \boldsymbol{\Sigma}_U + n\mathbf{D}_V \boldsymbol{\Sigma}_v \mathbf{D}_V$. The reason we multiply $\boldsymbol{\Sigma}_z$ by n is to be consistent with its sample estimates. Note that $\widehat{\boldsymbol{\Sigma}}_z = n^{-1}\sum_{i=1}^n \boldsymbol{z}_t \boldsymbol{z}_t^\intercal$ is an approximation of $\boldsymbol{\Sigma}_z$ while $\widehat{\boldsymbol{\Sigma}}_{x^*} = \sum_{i=1}^n (\boldsymbol{x}_t^*)(\boldsymbol{x}_t^*)^\intercal$ is an approximation of $\boldsymbol{\Sigma}_{x^*}$.

[1] The rules of using capital letters in the subscript are that (i) For factor loading matrices, magnitude matrices and number of factors, we use capital letters; (ii) For the integral volatility matrices of a continuouse processes, we use capital letters; (iii) For other cases such as the covariance matrices of discrete time series and the contaminated integral volatility matrices, e.g., $\boldsymbol{\Sigma}_{x^*}$ and $\boldsymbol{\Sigma}_z$, we choose according to the notation of the series.

Similar to the sparsity condition in Fan et al. (2013) and Aït-Sahalia and Xiu (2017), the following assumption is required in order to identify the factor structure. For ease of composition, we define

$$\bar{\tau}_G^{*+} = (1/2 + \bar{\tau}_G)_+, \quad \underline{\tau}_G^{*-} = (1/2 + \underline{\tau}_G)_-, \quad \bar{\tau}_V^{*} = (1/2 + \bar{\tau}_V), \quad \text{and } \bar{\tau}_V^{*+} = (1/2 + \bar{\tau}_V)_+.$$

***Assumption* 5.** *(Sparsity of idiosyncratic integrated covariance matrices) The integrated volatility matrices of the idiosyncratic components satisfy*

$$\|\mathbf{\Sigma}_U\|_1 = O(m_{U,d}), \quad \|\mathbf{\Sigma}_v\|_1 = O(m_{v,d}), \quad m_{w,nd}/(d^{1/2} n^{2\underline{\tau}_G^{*-}}) \to 0 \text{ with } m_{w,nd} = m_{U,d} + n^{2\bar{\tau}_V^*} m_{v,d}.$$

Under Assumption 5, we have $\|\mathbf{\Sigma}_w\|_1 = O(m_{w,nd})$. It is worth pointing out that the condition $m_{w,nd}/(dn^{2\underline{\tau}_G^{*-}}) \to 0$ is sufficient for the identification of the factors, see Lemma A.1. When only the weaker condition is required, we refer to the assumption as Assumption 5*. In order to obtain consistent estimates, we further require $m_{w,nd}/(d^{1/2} n^{2\underline{\tau}_G^{*-}}) \to 0$ as in Assumption 5. That estimation of an approximate factor model requires more strict sparsity conditions than identification is also observed in Aït-Sahalia and Xiu (2017).

We also make the following assumption on the stationarity of the microstructure noise components.

***Assumption* 6.** *(Stationary and sub-Gaussian microstructure noise)*
(i) The series $\{\boldsymbol{G}_t, \boldsymbol{V}_t\}$ *is strictly stationary. In addition,* $E[G_{jt}] = E[V_{it}] = E[G_{jt}V_{it}] = 0$ *for all* $1 \le i \le d$, $1 \le j \le K_G$ *and* $t = 0, \Delta, \ldots, n\Delta$.
(ii) There exist positive constants $C_\alpha > 0$ *and* $\gamma_1 > 0$ *such that the strong mixing condition of the series* $\{\boldsymbol{G}_t, \boldsymbol{V}_t\}$ *is satisfied with* $\alpha(s\Delta) \le C_\alpha \exp(-s^{\gamma_1})$.
(iii) There exist $b_1 > 0$, $\gamma_2 > 0$, *with* $\gamma_1^{-1} + 3\gamma_2^{-1} > 1$, *such that for all* $c > 0$, *we have*

$$\max_{1 \le j \le K_G} P\left(|G_{jt}| > c\right) \le \exp(1 - (c/b_1)^{\gamma_2}) \tag{2.8}$$

and

$$\max_{1 \le i \le d} P\left(|V_{it}| > c\right) \le \exp(1 - (c/b_1)^{\gamma_2}), \tag{2.9}$$

for $t = 0, \Delta, \ldots, n\Delta$.
(iv) There exist $b_2 > 0$, $\gamma_3 > 0$, *with* $\gamma_1^{-1} + 3\gamma_3^{-1} > 1$, *such that for all* $c > 0$, *we have*

$$\max_{1 \le j \le K_H} P\left(d^{-1/2} |\boldsymbol{\lambda}_{H,\cdot,j}^\intercal \boldsymbol{V}_t| > m_{v,d}^{1/2} \cdot c\right) \le \exp(1 - (c/b_2)^{\gamma_3}), \tag{2.10}$$

for $t = 0, \Delta, \ldots, n\Delta$, *where* $\boldsymbol{\lambda}_{H,\cdot,j}$ *is the* j*-th column of* $\mathbf{\Lambda}_H$*;*
(v) Let $1/\gamma = 1/\gamma_1 + 3/(\gamma_2 \vee \gamma_3)$. *Then,* $(\log d)^{2/\gamma - 1} = o(n)$.

Assumptions 6(i) and (ii) are similar to Assumption 3.2(i) and Assumption 3.3 in Fan et al. (2013). The strong mixing condition is more general than Assumptions 1 and 3 in Barigozzi et al. (2020b) in which the common components and idiosyncratic components of the microstructure noise are both assumed to be linear processes. The series $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$ can be serially correlated, which is more general than the assumption in Kim et al. (2016). Moreover, $\boldsymbol{V}_t$ can be cross-sectionally dependent, in which case, (2.2) gives an approximate factor model for the microstructure noise. More general assumptions can be found in Bai and Ng (2002) which permit weakly correlated idiosyncratic errors. Assumption 6(iii) requires exponential-type tails for the distributions of $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$, which is similar to Assumption 3.2(iii) of Fan et al. (2013) and Condition A1 of Tao et al. (2013b). If the microstructure noise has a heavier tail and the sub-Gaussian assumption is violated, robust estimation can be used to mitigate the influence of heavy tails (see, for example Fan and Kim (2018)). But to focus on the key objective of this paper (i.e., to identify and estimate factors for both efficient price and microstructure noise), we do not consider it in this paper.

Assumption 6(iv) is an additional exponential tail condition. Intuitively, let us consider two extreme cases. If $\boldsymbol{v}_t$ is cross-sectionally jointly independent, (2.10) holds true with $m_{v,d} = 1$. If $\boldsymbol{v}_t$ is cross-sectionally comonotonic, (2.10) holds true with $m_{v,d} = d$. The sparsity condition in Assumption 5 rules out the second case. But $\boldsymbol{v}_t$ can be cross-sectionally weakly dependent, with $m_{v,d}$ between 1 and $d^{1/2}n^{-1-2\tau_V+(1+2\tau_G)_-}$. Assumption 6(iv) guarantees that

$$\left\| (nd)^{-1}\sum_{s=1}^{n} \boldsymbol{\Lambda}_H^\intercal \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal \boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^\intercal \boldsymbol{\Sigma}_v \boldsymbol{\Lambda}_H \right\|_{\mathsf{MAX}} = O_P(m_{v,d}(\log d/n)^{1/2}),$$

see Lemma B.2(iv). Assumption 6(iv) can also be seen as an extension of the condition (see Assumption F.3 of Bai (2003) and Assumption 3.4(iii) of Fan et al. (2013))

$$\max_{1\le j\le K_H} \mathsf{E}[(d^{-1/2}|\boldsymbol{\lambda}_{H,\cdot,j}^\intercal \boldsymbol{v}_t|)^4] < C,$$

for some positive constant C. The latter condition is no longer true when $d^{-1}\|\boldsymbol{\Lambda}_H^\intercal \boldsymbol{\Sigma}_v \boldsymbol{\Lambda}_H\|_{\mathsf{MAX}} \to \infty$. Assumption 6(v) presents a trade-off between the mixing and tail conditions and the dimension d.

3. Estimation Procedure and Asymptotic Results

In this section, we develop a three-step estimation procedure for the common factors, $\mathscr{F}$ and $\mathscr{G}$, of the dual factor model. The estimation procedure is based on two PCA procedures and so we call it Double PCA or DPCA. The asymptotic results are presented step by step, so as to provide a better understanding on what the intermediate estimators estimate. In the first step, the factors and factor loadings for the combined factor model (2.4) in the first-difference form are estimated. In the second step, we cumulate the factors and separate the factors of the efficient prices from those of the microstructure noise in the first-difference form. In the last step, we cumulate $\widehat{\boldsymbol{f}}_t$ and $\widehat{\boldsymbol{g}}_t$ to get $\widehat{\boldsymbol{F}}_t$ and $\widehat{\boldsymbol{G}}_t$, respectively. For the time being, we assume that the number of factors K_F and K_G are known, and $K_H = K_F + K_G$. We will discuss how to determine them in Section 4.1.

An additional technical condition is required, which is similar to Assumption G in Bai (2003) and Assumption 4 in Aït-Sahalia and Xiu (2017) in the case of $\mathbf{D}_H = \mathbf{I}_{K_H}$. Note that we have assumed the boundedness of the largest eigenvalue of the matrix $d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Lambda}_H$ in Assumption 4, instead of the convergence of the matrix. Hence, we do not introduce a limiting matrix for it.

***Assumption* 7.** *(Distinct eigenvalues)*
The $d \times d$ matrix $\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Sigma}_h\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal$ has asymptotically distinct eigenvalues in the sense that

$$\frac{\mu_{j-1}(\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Sigma}_h\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal) - \mu_j(\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Sigma}_h\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal)}{\mu_{K_H}(\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Sigma}_h\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal)} > C,$$

when $(n, d) \to \infty$, for $j = 2, ..., K_H$, where C is a positive constant independent of n and d.

We also require the following assumption, which restricts the relation between n and d.

***Assumption* 8.** *(Relation between n and d) As $n \to \infty$,*
(i)

$$n^{1+4\underline{\tau}_G^{*-}-4(\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+})}/(\log d) \to \infty,$$

and
(ii)

$$n^{1+4\underline{\tau}_G^{*-}-4(\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+})}m_{w,nd}^2/(d^2\log d) \to 0.$$

When $\bar{\tau}_G = \underline{\tau}_G = \bar{\tau}_V = -1/2$ and $m_{w,nd} = O(1)$, Assumption 8 degenerates to $n/(\log d) \to \infty$ and $n/(d^2 \log d) \to 0$, which are similar to the corresponding condition in Theorem 3.1 of Fan et al. (2013) and assumption A1 of Tao et al. (2013b).

3.1. First step: PCA estimation in first-difference form

The first step in our estimation procedure is to apply PCA to the first-difference form of the dual factor model to extract estimates of all the factors and factor loadings (for both efficient prices and microstructure noise).

Recall that $\boldsymbol{x} = \boldsymbol{\hbar}\boldsymbol{\Lambda}_H^\intercal + \boldsymbol{w}$. Different identification conditions for the factor structure can be used, according to whether we would like to normalise the factors or the factor loadings. Accordingly, there are two ways to implement a PCA to estimate the factors and factor loadings. If we normalise the factor loadings, the PCA estimator solves the following optimisation problem,

$$\begin{cases} \min\limits_{\boldsymbol{\Lambda}_H,\boldsymbol{\hbar}} & \|\boldsymbol{x} - \boldsymbol{\hbar}\boldsymbol{\Lambda}_H^\intercal\|_F, \\ \mathsf{s.t.} & \boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Lambda}_H/d = \mathbf{I}_{K_H}. \end{cases}$$

Computationally, we conduct an eigen-decomposition of the $d \times d$ matrix $\boldsymbol{x}^\intercal \boldsymbol{x}$, and obtain $\widehat{\boldsymbol{\Lambda}}_H = \left(\widehat{\boldsymbol{\lambda}}_{H1}, \ldots, \widehat{\boldsymbol{\lambda}}_{Hd}\right)^\intercal$, which is the $d \times K_H$ matrix consisting of the K_H eigenvectors (multiplied by $\sqrt{d}$) corresponding to the K_H largest eigenvalues of $\boldsymbol{x}^\intercal \boldsymbol{x}$. The common factors can be estimated as

$$\widehat{\boldsymbol{\hbar}} = \boldsymbol{x}\widehat{\boldsymbol{\Lambda}}_H/d = \left(\widehat{\boldsymbol{h}}_\Delta, \ldots, \widehat{\boldsymbol{h}}_{n\Delta}\right)^\intercal. \tag{3.1}$$

Alternatively, if we normalise the factors, the PCA estimator solves the following optimisation problem,

$$\begin{cases} \min\limits_{\boldsymbol{\Lambda}_H, \boldsymbol{\hbar}} & \|\boldsymbol{x} - \boldsymbol{\hbar}\boldsymbol{\Lambda}_H^\intercal\|_F, \\ \mathsf{s.t.} & \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} = \mathbf{I}_{K_H}. \end{cases}$$

Computationally, we conduct PCA* (we add an asterisk to distinguish it from the PCA above that is based on the normalisation of the factor loadings) on the $n \times n$ matrix $\boldsymbol{x}\boldsymbol{x}^\intercal$, and get $\widehat{\boldsymbol{\hbar}}^* = \left(\widehat{\boldsymbol{h}}^*_\Delta, \ldots, \widehat{\boldsymbol{h}}^*_{n\Delta}\right)^\intercal$ denoting the $n \times K_H$ matrix consisting of the K_H eigenvectors corresponding to the K_H largest eigenvalues of $\boldsymbol{x}\boldsymbol{x}^\intercal$. The common factors can be estimated as

$$\widehat{\boldsymbol{\Lambda}}^*_H = \boldsymbol{x}^\intercal \widehat{\boldsymbol{\hbar}}^* = \left(\widehat{\boldsymbol{\lambda}}^*_{H1}, \ldots, \widehat{\boldsymbol{\lambda}}^*_{Hd}\right)^\intercal. \tag{3.2}$$

It is easy to show that both $\widehat{\boldsymbol{\hbar}}$ and $\widehat{\boldsymbol{\hbar}}^*$ are eigenvectors of $\boldsymbol{x}\boldsymbol{x}^\intercal$, and thereby, they span the same space. In this sense, the two ways of PCA are equivalent. Bai and Li (2012) point out that the analysis of one PCA representation will carry over to the other by switching the roles of n and d and the role of the factor loadings and factors, and thus it is sufficient to carefully examine the asymptotic properties of one representation. All high-frequency factor models, as we know, use (**??**) to implement PCA. The main reason may be that the spot covariance matrices is time-varying and thereby the $n \times n$ matrix $\boldsymbol{x}\boldsymbol{x}^\intercal$ is conceptually more difficult to analyse.

However, in our three-step estimation procedure, the final estimators based on the above two PCAs will not be equivalent. Since the first-step factor estimators will be fed into the second step in cumulative form for another PCA, whether the factors are normalised in the first step will affect the final results. In Section 3.3, we will compare the small sample performance of estimators based on the two different PCA representations in the first step. We will see that PCA* always outperforms PCA. Hence, we will use PCA* for our first step and derive the asymptotic theory based on it.

Since the asymptotic theory of $\widehat{\boldsymbol{\hbar}}$ is easier to establish than that of $\widehat{\boldsymbol{\hbar}}^*$, as discussed above, we first prove the consistency of $\widehat{\boldsymbol{\hbar}}$ (see Lemma A.2 in Appendix) and then use the following relation

$$\widehat{\boldsymbol{\hbar}}^* = \widehat{\boldsymbol{\hbar}}(\widehat{\boldsymbol{\hbar}}^\intercal \widehat{\boldsymbol{\hbar}})^{-1/2} \quad \text{and} \quad \widehat{\boldsymbol{\Lambda}}^*_H = \widehat{\boldsymbol{\Lambda}}_H(\widehat{\boldsymbol{\hbar}}^\intercal \widehat{\boldsymbol{\hbar}})^{1/2}, \tag{3.3}$$

to prove the consistency of $\widehat{\hbar}^*$. When d is much larger than n, it is computationally more convenient to conduct PCA on the $n \times n$ matrix $\boldsymbol{x}\boldsymbol{x}^\intercal$, and vice versa. Then the relations in (3.3) can be used to get the desired form of estimates.

The following theorem shows the uniform rate of convergence for $\widehat{\boldsymbol{\Lambda}}^*_H$ and $\widehat{\hbar}^*$ of the dual factor model. For ease of composition, we denote

$$a_{nd} = (\log d)^{1/2} \frac{n^{\bar{\tau}_V^{*+} + \bar{\tau}_G^{*+} \vee \bar{\tau}_V^{*+}}}{n^{1/2}} + \frac{m_{w,nd}}{d^{1/2}}.$$

.

Theorem 3.1. *Suppose that Assumptions 1–8 are satisfied. We have*
(i)

$$\left\| \widehat{\hbar}^{*\intercal} - (\mathbf{R}^*)^{-1} (\hbar^\intercal \hbar)^{-1/2} \hbar^\intercal \right\|_{\mathsf{MAX}} \leq \left\| \widehat{\hbar}^{*\intercal} - (\mathbf{R}^*)^{-1} (\hbar^\intercal \hbar)^{-1/2} \hbar^\intercal \right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right), \tag{3.4}$$

where the rotation matrix $\mathbf{R}^*$ is defined by

$$\mathbf{R}^* = d^{1/2} (\hbar^\intercal \hbar)^{-1/2} \boldsymbol{\Lambda}_H^\intercal \widehat{\boldsymbol{\Lambda}}_H \widehat{\mathbf{D}}_{x,K_H}^{-1/2},$$

in which $\widehat{\mathbf{D}}_{x,K_H} = d\widehat{\hbar}^\intercal \widehat{\hbar} = \widehat{\boldsymbol{\Lambda}}_H^{*\intercal} \widehat{\boldsymbol{\Lambda}}_H^*$ is a $K_H \times K_H$ diagonal matrix with the diagonal elements being the first K_H largest eigenvalues of $\boldsymbol{x}^\intercal \boldsymbol{x}$ arranged in a descending order;
(ii)

$$\left\| \widehat{\boldsymbol{\Lambda}}_H^* - \boldsymbol{\Lambda}_H \mathbf{D}_H (\hbar^\intercal \hbar)^{1/2} \mathbf{R}^* \right\|_{\mathsf{MAX}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right); \tag{3.5}$$

(iii) $\widehat{\hbar}^* \widehat{\boldsymbol{\Lambda}}_H^{*\intercal} = \widehat{\hbar} \widehat{\boldsymbol{\Lambda}}_H^\intercal$ and

$$\left\| \widehat{\hbar}^* \widehat{\boldsymbol{\Lambda}}_H^{*\intercal} - \hbar \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal \right\|_{\mathsf{MAX}} = O_P\left(n^{\bar{\tau}_G^{*+} - 2\underline{\tau}_G^{*-}} \cdot a_{nd} \right); \tag{3.6}$$

(iv) $\mathbf{R}^*$ is an asymptotically orthogonal matrix, that is

$$\| \mathbf{R}^{*\intercal} \mathbf{R}^* - \mathbf{I}_{K_H} \|_{\mathsf{sp}} = O_P(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd}).$$

Note that $n^{-2\underline{\tau}_G^{*-}} a_{nd} = o(1)$ under Assumption 8. The theorem shows the uniform rate of convergence for $\widehat{\hbar}^*$ of the dual factor model, which is similar to Theorem 5 in Aït-Sahalia and Xiu (2017) for high-frequency factor model without the magnitude matrices. The estimator $\widehat{\hbar}^*$ converges to normalised factors $(\hbar^\intercal \hbar)^{-1/2} \hbar^\intercal$ up to an asymptotically orthogonal matrix. The result is different from the uniform results for low-frequency factor models, e.g., Proposition 2 in Bai (2003) and Theorem 3.3 in Fan et al. (2013). In a low-frequency factor model, the uniform convergence of the estimator for $n^{1/2}(\boldsymbol{\hbar}^\intercal \boldsymbol{\hbar})^{-1/2}\boldsymbol{\hbar}^\intercal$ can be derived, while in a high-frequency factor model, only the uniform convergence of the estimator for $(\boldsymbol{\hbar}^\intercal \boldsymbol{\hbar})^{-1/2}\boldsymbol{\hbar}^\intercal$ can be achieved.

The introduction of the magnitude matrices provides an insight into how larger noise-to-signal ratio can worsen the estimation. We can see that the convergence rates are affected by the magnitudes of both $\mathbf{D}_G$ and $\mathbf{D}_V$, by noticing that $\underline{\tau}_G^{*-}$ directly affects the convergence rate and that $\bar{\tau}_G^{*+}$ and $\bar{\tau}_V^{*+}$ affect a_{nd}. The convergence rates are fastest (i.e., $(\log d)^{1/2}n^{-1/2} + m_{w,nd}d^{-1/2}$) when $\bar{\tau}_G = \underline{\tau}_G = \bar{\tau}_V = -1/2$. When common components of the microstructure noise have a larger or smaller magnitude than that of the efficient prices, the estimators will be worse. Note that $n^{\bar{\tau}_G^{*+} - 2\underline{\tau}_G^{*-}} \cdot a_{nd}$ (see Theorem 3.1 (iii)) grows with $\bar{\tau}_G$ when $\bar{\tau}_G > -1/2$, and thus it may not converge to zero. However, this is not a big problem, since the convergence of the estimator of common components still holds if we re-scale the data by a factor of order $n^{-\bar{\tau}_G^{*+}}$. Nevertheless, if we require $n^{\bar{\tau}_G^{*+} - 2\underline{\tau}_G^{*-}} \cdot a_{nd} = o(1)$, we only need to extend Assumption 8(i) to $n^{1+4\underline{\tau}_G^{*-} - 2\bar{\tau}_G^{*+} - 2\bar{\tau}_V^{*+} - 2(\bar{\tau}_G^{*+} \vee \bar{\tau}_V^{*+})}/(\log d) \to \infty$.

3.2. Second step: PCA estimation in cumulative form

The second step aims to obtain a $K_H \times K_G$ matrix $\widehat{\boldsymbol{\beta}}$ such that $\widehat{\boldsymbol{\hbar}}^*\widehat{\boldsymbol{\beta}}$ estimates the factors for microstructure noise. Denote $\boldsymbol{\beta} = (\mathbf{O}_{K_G \times K_F} \quad \mathbf{I}_{K_G})^\intercal$ and $\boldsymbol{\beta}_\perp = (\mathbf{I}_{K_F} \quad \mathbf{O}_{K_F \times K_G})^\intercal$. Then $\boldsymbol{g} := \boldsymbol{\hbar}\boldsymbol{\beta}$ is the matrix of true factors for the difference of microstructure noise, and $\boldsymbol{f} := \boldsymbol{\hbar}\boldsymbol{\beta}_\perp$ is the matrix of true factors for the difference of efficient prices. However, due to rotational indeterminacy of the estimated factors, $\widehat{\boldsymbol{\hbar}}^*$, instead of applying $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$ directly on $\widehat{\boldsymbol{\hbar}}^*$ to obtain estimates of $\boldsymbol{g}$ and $\boldsymbol{f}$, we need to find proper rotations of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$ that achieve this goal.

There are different ways to estimate such rotations of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$. For example, Barigozzi et al. (2020b) use Johansen (1995)'s reduced rank estimation in a dynamic factor model to estimate the cointegration coefficients of non-stationary factors (also see Section 4.2). However, their method requires the specification of a finite-order vector autoregression (or a vector error correction model) prior to estimation. By contrast, we will use a second-step PCA on cumulated factors to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$, which allows for more general (non)stationarity structure of the underlying processes. This second step is similar to PANIC due to Bai and Ng (2004), which is developed from Stock and Watson (1988) and Harris (1997).

For $1 \leq s \leq n$, let $\widehat{\boldsymbol{H}}_{s\Delta}^* = \sum_{s_1=1}^{s} \widehat{\boldsymbol{h}}_{s_1\Delta}^*$ be an estimator of $\boldsymbol{H}_{s\Delta}$. Define the demeaned $\widehat{\boldsymbol{H}}_{s\Delta}^*$ as $\widehat{\boldsymbol{H}}_{s\Delta}^{*c} = \widehat{\boldsymbol{H}}_{s\Delta}^* - \overline{\widehat{\boldsymbol{H}}^*}$, where $\overline{\widehat{\boldsymbol{H}}^*} = n^{-1}\sum_{s=1}^{n} \widehat{\boldsymbol{H}}_{s\Delta}^*$. In the matrix form, this can be written as $\widehat{\mathscr{H}}^{*c} = \widehat{\mathscr{H}}^* - \overline{\widehat{\mathscr{H}}^*}$, with $\widehat{\mathscr{H}}^{*c} = (\widehat{\boldsymbol{H}}_{\Delta}^{*c}, \ldots, \widehat{\boldsymbol{H}}_{n\Delta}^{*c})^\intercal$, $\widehat{\mathscr{H}}^* = (\widehat{\boldsymbol{H}}_{\Delta}^*, \ldots, \widehat{\boldsymbol{H}}_{n\Delta}^*)^\intercal$, and $\overline{\widehat{\mathscr{H}}^*} = \mathbf{1}_n \overline{\widehat{\boldsymbol{H}}^*}^\intercal$. Let $\widehat{\boldsymbol{\beta}}_\perp$ be the matrix of eigenvectors associated with the largest K_F eigenvalues of $n^{-1}\widehat{\mathscr{H}}^{*c\intercal}\widehat{\mathscr{H}}^{*c}$ and $\widehat{\boldsymbol{\beta}}$ be the matrix of eigenvectors associated with the rest of the K_G eigenvalues. Then, $\widehat{\boldsymbol{\beta}}_\perp^\intercal \widehat{\boldsymbol{h}}_t^*$ is an estimator of $\boldsymbol{\beta}_\perp^\intercal \boldsymbol{h}_t = \boldsymbol{f}_t$, and $\widehat{\boldsymbol{\beta}}^\intercal \widehat{\boldsymbol{h}}_t^*$ is an estimator of $\boldsymbol{\beta}^\intercal \boldsymbol{h}_t = \boldsymbol{g}_t$.

Lemma 3.1. *Under the assumptions of Theorem 3.1, $\widehat{\boldsymbol{\beta}}_\perp$ and $\widehat{\boldsymbol{\beta}}$ are super-consistent in the sense that*

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\Xi}^\intercal \boldsymbol{\beta} \mathbf{Q}_{\boldsymbol{\beta}} = O_P(n^{-1}) \tag{3.7}$$

and

$$\widehat{\boldsymbol{\beta}}_\perp - \boldsymbol{\Xi}^{-1} \boldsymbol{\beta}_\perp \mathbf{Q}_{\boldsymbol{\beta}_\perp} = O_P(n^{-1}). \tag{3.8}$$

where

$$\mathbf{Q}_{\boldsymbol{\beta}} = [\boldsymbol{\beta}^\intercal \boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal \boldsymbol{\beta}]^{-1} \boldsymbol{\beta}^\intercal \boldsymbol{\Xi} \widehat{\boldsymbol{\beta}}, \qquad \mathbf{Q}_{\boldsymbol{\beta}_\perp} = [\boldsymbol{\beta}_\perp^\intercal (\boldsymbol{\Xi}^\intercal)^{-1} \boldsymbol{\Xi}^{-1} \boldsymbol{\beta}_\perp]^{-1} \boldsymbol{\beta}_\perp^\intercal (\boldsymbol{\Xi}^\intercal)^{-1} \widehat{\boldsymbol{\beta}}_\perp,$$
$$\boldsymbol{\Xi} = d^{1/2} \boldsymbol{\Lambda}_H^\intercal \widehat{\boldsymbol{\Lambda}}_H \widehat{\mathbf{D}}_{x,K_H}^{-1/2} = (\boldsymbol{\hbar}^\intercal \boldsymbol{\hbar})^{1/2} \mathbf{R}^*.$$

The lemma shows that $\widehat{\boldsymbol{\beta}}$ estimates a basis for the space spanned by $\boldsymbol{\Xi}^\intercal \boldsymbol{\beta}$. Using Lemma 3.1 and Theorem 3.1, we can prove the following theorem, which gives the convergence rate for the estimators of the factors for microstructure noise and the factor loadings for efficient prices. To this end, we define

$$\widehat{\boldsymbol{f}}^* = \widehat{\boldsymbol{\hbar}}^* \widehat{\boldsymbol{\beta}}_\perp, \quad \widehat{\boldsymbol{g}}^* = \widehat{\boldsymbol{\hbar}}^* \widehat{\boldsymbol{\beta}}, \quad \widehat{\boldsymbol{\Lambda}}_F^* = \widehat{\boldsymbol{\Lambda}}_H^* \widehat{\boldsymbol{\beta}}_\perp, \quad \text{and} \quad \widehat{\boldsymbol{\Lambda}}_G^* = \widehat{\boldsymbol{\Lambda}}_H^* \widehat{\boldsymbol{\beta}}.$$

Theorem 3.2. *Suppose that Assumptions 1–8 are satisfied. We have*

(i)

$$\left\| \widehat{\boldsymbol{f}}^* - \boldsymbol{f} \boldsymbol{\beta}_\perp^\intercal (\boldsymbol{\Xi}^\intercal)^{-1} \widehat{\boldsymbol{\beta}}_\perp \right\|_{\mathsf{MAX}} \le \left\| \widehat{\boldsymbol{f}}^* - \boldsymbol{f} \boldsymbol{\beta}_\perp^\intercal (\boldsymbol{\Xi}^\intercal)^{-1} \widehat{\boldsymbol{\beta}}_\perp \right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right); \tag{3.9}$$

(ii)

$$\| \widehat{\boldsymbol{g}}^* - \boldsymbol{g} \mathbf{Q}_{\boldsymbol{\beta}} \|_{\mathsf{MAX}} \le \| \widehat{\boldsymbol{g}}^* - \boldsymbol{g} \mathbf{Q}_{\boldsymbol{\beta}} \|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right); \tag{3.10}$$

(iii)

$$\left\| \widehat{\boldsymbol{\Lambda}}_F^* - \boldsymbol{\Lambda}_F \mathbf{Q}_{\boldsymbol{\beta}_\perp} \right\|_{\mathsf{MAX}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right); \tag{3.11}$$

(iv)

$$\left\| \widehat{\boldsymbol{\Lambda}}_G^* - \boldsymbol{\Lambda}_G \mathbf{D}_G \boldsymbol{\beta}^\intercal \boldsymbol{\Xi} \widehat{\boldsymbol{\beta}} \right\|_{\mathsf{MAX}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right), \tag{3.12}$$

where $\mathbf{Q}_{\boldsymbol{\beta}}$, $\mathbf{Q}_{\boldsymbol{\beta}_\perp}$ and $\boldsymbol{\Xi}$ are defined in Lemma 3.1.

This lemma shows that $\widehat{\boldsymbol{f}}^*$ and $\widehat{\boldsymbol{g}}^*$ are estimators of the factors for the first-differenced efficient prices and microstructure noise with rotations $\boldsymbol{\beta}_\perp^\intercal (\boldsymbol{\Xi}^\intercal)^{-1} \widehat{\boldsymbol{\beta}}_\perp$ and $\mathbf{Q}_{\boldsymbol{\beta}}$, respectively, and that $\widehat{\boldsymbol{\Lambda}}_F^*$ and $\widehat{\boldsymbol{\Lambda}}_G^*$ are estimators of the factor loadings for efficient prices and microstructure noise with rotations $\mathbf{Q}_{\boldsymbol{\beta}_\perp}$ and $\boldsymbol{\beta}^\intercal \boldsymbol{\Xi} \widehat{\boldsymbol{\beta}}$, respectively. When we estimate the first-differenced common components for microstructure noise and efficient prices (i.e., $\boldsymbol{g} \boldsymbol{\Lambda}_G^\intercal$ and $\boldsymbol{f} \boldsymbol{\Lambda}_F^\intercal$), these rotations cancel out. Thus, we have the following corollary.

Corollary 3.1. *Suppose that Assumptions 1–8 are satisfied. We have*
(i)

$$\left\| \widehat{\mathscr{f}}^* \widehat{\mathbf{\Lambda}}_F^{*\intercal} - \mathscr{f} \mathbf{\Lambda}_F^\intercal \right\|_{\mathsf{MAX}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right); \tag{3.13}$$

(ii)

$$\left\| \widehat{\mathscr{g}}^* \widehat{\mathbf{\Lambda}}_G^{*\intercal} - \mathscr{g} \mathbf{D}_G \mathbf{\Lambda}_G^\intercal \right\|_{\mathsf{MAX}} = O_P\left(n^{\bar{\tau}_G^* - 2\underline{\tau}_G^{*-}} \cdot a_{nd} \right), \tag{3.14}$$

where $\bar{\tau}_G^* = 1/2 + \bar{\tau}_G$ *and* $\underline{\tau}_G^{*-} = (1/2 + \underline{\tau}_G)_-$.

3.3. Third step: Cumulation of factors

Now we construct the estimates of $\mathscr{F}$ and $\mathscr{G}$ by cumulating $\widehat{\mathscr{f}}^* = \left(\widehat{\boldsymbol{f}}_\Delta^*, \ldots, \widehat{\boldsymbol{f}}_{n\Delta}^* \right)^\intercal$ and $\widehat{\mathscr{g}}^* = \left(\widehat{\boldsymbol{g}}_\Delta^*, \ldots, \widehat{\boldsymbol{g}}_{n\Delta}^* \right)^\intercal$, respectively. That is, $\widehat{\mathscr{F}}^* = \left(\widehat{\boldsymbol{F}}_\Delta^*, \ldots, \widehat{\boldsymbol{F}}_{n\Delta}^* \right)^\intercal$ and $\widehat{\mathscr{G}}^* = \left(\widehat{\boldsymbol{G}}_\Delta^*, \ldots, \widehat{\boldsymbol{G}}_{n\Delta}^* \right)^\intercal$, where

$$\widehat{\boldsymbol{F}}_t^* = \sum_{s:0 \le s\Delta \le t} \widehat{\boldsymbol{f}}_{s\Delta}^* \quad \text{and} \quad \widehat{\boldsymbol{G}}_t^* = \sum_{s:0 \le s\Delta \le t} \widehat{\boldsymbol{g}}_{s\Delta}^*.$$

From the second step, we also obtain $\widehat{\mathscr{H}^*} = \left(\widehat{\boldsymbol{H}}_\Delta^*, \ldots, \widehat{\boldsymbol{H}}_{n\Delta}^* \right)^\intercal$, where $\widehat{\boldsymbol{H}}_t^* = \sum_{s:0 \le s\Delta \le t} \widehat{\boldsymbol{h}}_{s\Delta}^*$. More compactly, we can write $\mathscr{F} = \mathbf{L}_n \mathscr{f}$, $\mathscr{G} = \mathbf{L}_n \mathscr{g}$ and $\mathscr{H} = \mathbf{L}_n \mathscr{h}$, and their estimates, $\widehat{\mathscr{F}}^* = \mathbf{L}_n \widehat{\mathscr{f}}^*$, $\widehat{\mathscr{G}}^* = \mathbf{L}_n \widehat{\mathscr{g}}^*$ and $\widehat{\mathscr{H}^*} = \mathbf{L}_n \widehat{\mathscr{h}}^*$. Since $\|\mathbf{L}_n\|_{\mathsf{sp}} = 1$, the convergence rates for the estimators of the first-differenced factors are also the convergence rates for the estimators for the aggregated factors. This leads to the following theorem.

Theorem 3.3. *Suppose that Assumptions 1–8 are satisfied. Then,*
(i) for factors, we have

$$\left\| \widehat{\mathscr{H}^*} - (\mathscr{H} - \mathbf{1}_n \boldsymbol{H}_0^\intercal)(\mathbf{\Xi}^\intercal)^{-1} \right\|_{\mathsf{MAX}} \le \left\| \widehat{\mathscr{H}^*} - (\mathscr{H} - \mathbf{1}_n \boldsymbol{h}_0^\intercal)(\mathbf{\Xi}^\intercal)^{-1} \right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right), \tag{3.15}$$

$$\left\| \widehat{\mathscr{G}}^* - (\mathscr{G} - \mathbf{1}_n \boldsymbol{G}_0^\intercal)\mathbf{Q}_{\boldsymbol{\beta}} \right\|_{\mathsf{MAX}} \le \left\| \widehat{\mathscr{G}}^* - (\mathscr{G} - \mathbf{1}_n \boldsymbol{G}_0^\intercal)\mathbf{Q}_{\boldsymbol{\beta}} \right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right), \tag{3.16}$$

and

$$\begin{aligned} \left\| \widehat{\mathscr{F}}^* - (\mathscr{F} - \mathbf{1}_n \boldsymbol{F}_0^\intercal)\boldsymbol{\beta}_\perp^\intercal (\mathbf{\Xi}^\intercal)^{-1} \widehat{\boldsymbol{\beta}}_\perp \right\|_{\mathsf{MAX}} &\le \left\| \widehat{\mathscr{F}}^* - (\mathscr{F} - \mathbf{1}_n \boldsymbol{F}_0^\intercal)\boldsymbol{\beta}_\perp^\intercal (\mathbf{\Xi}^\intercal)^{-1} \widehat{\boldsymbol{\beta}}_\perp \right\|_{\mathsf{sp}} \\ &= O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd} \right); \end{aligned} \tag{3.17}$$

(ii) for common components, we have

$$\left\|\widehat{\mathscr{H}^*}(\widehat{\mathbf{\Lambda}}_H^*)^\intercal - (\mathscr{H} - \mathbf{1}_n \boldsymbol{H}_0^\intercal)\mathbf{D}_H \mathbf{\Lambda}_H^\intercal\right\|_{\mathsf{MAX}} = O_P\left(n^{\bar{\tau}_G^{*+} - 2\underline{\tau}_G^{*-}} \cdot a_{nd}\right), \tag{3.18}$$

$$\left\|\widehat{\mathscr{G}}^*(\widehat{\mathbf{\Lambda}}_G^*)^\intercal - (\mathscr{G} - \mathbf{1}_n \boldsymbol{G}_0^\intercal)\mathbf{D}_G \mathbf{\Lambda}_G^\intercal\right\|_{\mathsf{MAX}} = O_P\left(n^{\bar{\tau}_G^{*} - 2\underline{\tau}_G^{*-}} \cdot a_{nd}\right), \tag{3.19}$$

and

$$\left\|\widehat{\mathscr{F}}^*(\widehat{\mathbf{\Lambda}}_F^*)^\intercal - (\mathscr{F} - \mathbf{1}_n \boldsymbol{F}_0^\intercal)\mathbf{\Lambda}_F^\intercal\right\|_{\mathsf{MAX}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right), \tag{3.20}$$

where $\mathbf{Q}_{\boldsymbol{\beta}}$, $\mathbf{Q}_{\boldsymbol{\beta}_\perp}$ and $\mathbf{\Xi}$ are defined in Lemma 3.1.

When $\bar{\tau}_G = \underline{\tau}_G = \bar{\tau}_V = -1/2$ and $m_{w,nd} = O(1)$, the uniform convergence rate of $\widehat{\mathscr{H}^*}$ is $O_P((\log d)^{1/2} n^{-1/2} + d^{-1/2})$. In comparison, the uniform consistency result for the low-frequency factor model in Bai and Ng (2004) is $O_P\left(n^{-3/4} + d^{-1/2}\right)$ (see Lemma 2 in Bai and Ng (2004)). Our estimators have slower convergence rates, which is mainly due to the different settings.

4. Simulation

4.1. Number of factors

In this paper, we apply the commonly-used information criterion proposed by Bai and Ng (2002) to estimate the total number of factors. Then, we use the PANIC test procedure proposed by Bai and Ng (2004) to determine the number of factors for efficient prices, and compare it to Hallin and Liška (2007)'s spectral method.

To introduce Bai and Ng (2002)'s information criterion, we denote by $\overline{Q}$ a finite positive integer that is no smaller than K_H. For any $1 \le q_H \le \overline{Q}$, we let $\widehat{\mathscr{h}}^*(q_H) = \left(\widehat{\boldsymbol{h}}_\Delta^*(q_H), \dots, \widehat{\boldsymbol{h}}_{n\Delta}^*(q_H)\right)^\intercal$ be the matrix of estimated factors when the total number of factors is assumed to be q_H, and denote by $\widehat{\mathbf{\Lambda}}_H^*(q_H)$ the corresponding loadings matrix. The information criterion is defined as

$$\mathsf{IC}_1(q_H) = \log\left[\mathsf{V}_n(q_H)\right] + q_H\left(\frac{n+d}{nd}\right)\log(\frac{nd}{n+d}), \tag{4.1}$$

where $\mathsf{V}_n(q_H) = \|\boldsymbol{x} - \widehat{\mathscr{h}}^*(q_H)(\widehat{\mathbf{\Lambda}}_H^*(q_H))^\intercal\|_F$. The total number of factors is then estimated as

$$\widehat{K}_H = \underset{0 \le q_H \le \overline{Q}}{\arg\min}\, \mathsf{IC}_1(q_H), \tag{4.2}$$

with $\mathsf{IC}_1(0) = \|\boldsymbol{x}\|_F$ for convention. For consistency of $\widehat{K}_H$, we refer to the asymptotic results given in Theorem 2 of Bai and Ng (2002).

Bai and Ng (2004) propose two tests to determine the number of nonstationary factors. We adopt the one that does not specify a finite order VAR representation. For any $1 \le q_F \le \widehat{K}_H$, we let $\widehat{\mathscr{F}}^*(q_F) = \widehat{\mathscr{H}}^* \widehat{\boldsymbol{\beta}}_\perp(q_F)$, when the number of factors for efficient prices is assumed to be q_F. Let $\widehat{\xi}_t^F$ be the residuals from estimating a first-order VAR of $\widehat{\mathscr{F}}^*(q_F)$ and $\widehat{\boldsymbol{\Sigma}}_{L,F}(q_F)$ be the estimated long-run covariance matrix of $\widehat{\xi}_t^F$. The test statistic for $H_0 : K_F = q_F$ is defined by

$$MQ(q_F) = n(\nu(q_F) - 1) \tag{4.3}$$

where $\nu(q_F)$ is the smallest eigenvalue of

$$\left[\sum_{s=2}^{n} \frac{1}{2}\left(\widehat{\boldsymbol{F}}^*_{s\Delta}(q_F)\widehat{\boldsymbol{F}}^{*\intercal}_{(s-1)\Delta}(q_F) + \widehat{\boldsymbol{F}}^*_{(s-1)\Delta}(q_F)\widehat{\boldsymbol{F}}^{*\intercal}_{s\Delta}(q_F)\right) - n\widehat{\boldsymbol{\Sigma}}_{L,F}(q_F)\right]\Big(\sum_{s=2}^{n} \widehat{\boldsymbol{F}}^*_{s\Delta}(q_F)\widehat{\boldsymbol{F}}^{*\intercal}_{s\Delta}(q_F)\Big)^{-1}. \tag{4.4}$$

For a given significance level α, define

$$\widehat{K}_F = \max_{0 \le q_F \le \widehat{K}_H,\ p_{MQ}(q_F) > \alpha} q_F, \tag{4.5}$$

where $p_{MQ}(q_F)$ is the p-value of the statistic $MQ(q_F)$ obtained by simulation using vector standard Brownian motions, and we define $p_{MQ}(0) = 1$ for convention. We refer the reader to Theorem 1 of Bai and Ng (2004) for the asymptotic distribution of the test statistic.

Hallin and Liška (2007)'s spectral method can also be used to estimate the number of factors for efficient prices. Let $\widehat{\Gamma}_k$ be the $d \times d$ sample lag-k autocovariance matrix of $\boldsymbol{x}_t$. Define the lag window estimator of the spectral density matrix of $\boldsymbol{x}_t$ by

$$\widehat{\boldsymbol{\Sigma}}_x(\theta) = \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} \widehat{\Gamma}_k e^{-ik\theta} w(B_n^{-1} k), \tag{4.6}$$

where B_n is a suitable bandwidth and $w(\cdot)$ is a positive even weight function. Let $\nu_l(\theta)$ be the l-th largest eigenvalue of $\widehat{\boldsymbol{\Sigma}}_x(\theta)$ and define the following information criteria

$$\mathsf{IC}_{2.i}(q) = \log\left(\frac{1}{n(2B_n+1)} \sum_{h=-B_n}^{B_n} \sum_{l=q+1}^{d} \nu_l(\theta_h)\right) + q s_i(n,d), \quad i = 1,2,3, \tag{4.7}$$

and

$$\mathsf{IC}_{3.i}(q) = \log\left(\frac{1}{n} \sum_{l=q+1}^{d} \nu_l(0)\right) + q s_i(n,d), \quad i = 1,2,3, \tag{4.8}$$

where, with $M_n = \lfloor 0.75\sqrt{n} \rfloor$ and a constant ι,

$$s_1(n,d) = \iota \cdot (M_n^{-2} + M_n^{1/2} n^{-1/2} + d^{-1}) \cdot \log(\min\{M_n^2, M_n^{-1/2} n^{1/2}, d\}),$$

$$s_2(n,d) = \iota \cdot (\min\{M_n^2, M_n^{-1/2} n^{1/2}, d\})^{-1/2},$$

and

$$s_3(n,d) = \iota \cdot (\min\{M_n^2, M_n^{-1/2} n^{1/2}, d\})^{-1} \cdot \log(\min\{M_n^2, M_n^{-1/2} n^{1/2}, d\}).$$

Barigozzi et al. (2020b) show that K_H can be estimated as

$$\widehat{K}_H = \arg\min_{1 \le q \le \overline{Q}} \mathsf{IC}_{2.i}(q) \tag{4.9}$$

for any $i = 1, 2, 3$, and K_F can be estimated as

$$\widehat{K}_F = \arg\min_{1 \le q \le \overline{Q}} \mathsf{IC}_{3.i}(q) \tag{4.10}$$

for any $i = 1, 2, 3$. In practice, one can let $\overline{Q} = \widehat{K}_H$ in (4.10) to make sure $\widehat{K}_F \le \widehat{K}_H$.

4.2. Alternative approaches for comparison

We consider two alternative approaches for comparison. The first approach, denoted as DPCA, estimates the factors as $\widehat{\hbar}$ in (3.1) instead of $\widehat{\hbar}^*$, and uses $\widehat{\hbar}$ for the second-step PCA while keeping the rest of the steps exactly the same. The second approach is proposed by Barigozzi et al. (2020a) and Barigozzi et al. (2020b) and is denoted as PCA*-VECM. The PCA*-VECM uses $\widehat{\boldsymbol{h}}_t^*$ from the first-step PCA* to construct a Vector Error Correction Model (VECM) in order to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$, while in our method, we use a second-step PCA to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$. We set the lag of VECM to 1 for simplicity, and for a lag length larger than 1, we refer the reader to Chapter 6 of Johansen (1995). Specifically, the PCA*-VECM uses Johansen (1995)'s reduced rank regression method to estimate a VECM for $\widehat{\boldsymbol{h}}_t^*$:

Step 1: Implement OLS of $\widehat{\boldsymbol{h}}_t^*$ and $\widehat{\boldsymbol{H}}_{t-\Delta}^*$ on $\widehat{\boldsymbol{h}}_{t-\Delta}^*$ to get residuals $\widehat{\boldsymbol{e}}_{0,t}$ and $\widehat{\boldsymbol{e}}_{1,t}$, respectively.

Step 2: Let $\widehat{\mathbf{S}}_{ij} = n^{-1}\sum_{s=1}^{n} \widehat{\boldsymbol{e}}_{i,s\Delta}\widehat{\boldsymbol{e}}_{j,s\Delta}^\intercal$ for $i, j = 0, 1$. Then let $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_{K_G})$, where $\widehat{\boldsymbol{\beta}}_l$, $l = 1, \ldots, K_G$, is the eigenvector belonging to the l-th largest eigenvalue of the matrix $(\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01})$. Define $\widehat{\boldsymbol{\beta}}_\perp$ as the orthogonal complement matrix of $\widehat{\boldsymbol{\beta}}$ such that $\widehat{\boldsymbol{\beta}}_\perp^\intercal \widehat{\boldsymbol{\beta}} = \mathbf{O}_{K_F \times K_G}$ and $\widehat{\boldsymbol{\beta}}_\perp^\intercal \widehat{\boldsymbol{\beta}}_\perp = \mathbf{I}_{K_F}$.

Note that even if the estimated factors are not normalised in Step 1, the residuals $\widehat{\boldsymbol{e}}_{0,t}$ and $\widehat{\boldsymbol{e}}_{1,t}$ will not change and therefore the estimates of $\widehat{\boldsymbol{\beta}}_\perp$ and $\widehat{\boldsymbol{\beta}}$ will not be affected. Also note that the factors for efficient prices follow a diffusion model and hence, are heteroskedastic. One can implement more efficient estimation of VECM under heteroskedasticity (e.g., generalized least squares estimation in Seo (2007) and Herwartz and Lütkepohl (2011)). But we do not pursue this in our paper.

4.3. Data generating processes

The data generating process in our simulation is similar to that in Bollerslev et al. (2019). The observable prices are the sum of efficient prices and microstructure noise. The former has two orthogonal factors and the latter has one factor.

The two factors in efficient prices are independently generated from a GARCH diffusion model as in Andersen and Bollerslev (1998),

$$\begin{cases} f_{it} = \sigma_{fit} \mathsf{d} B_{it}, \\ \mathsf{d}\sigma^2_{fit} = \kappa_{fi}(\theta_{fi} - \sigma^2_{fit})\mathsf{d}t + \lambda_{fi}\sigma^2_{fit}\mathsf{d}W_{it}, \end{cases} \tag{4.11}$$

for $i = 1, 2$, where B_{it} and W_{it} are dependent Brownian motions with $\mathsf{corr}(B_{it}, W_{it}) = -0.5$. The parameters are set as $\kappa_{f1} = \kappa_{f2} = 0.035$, $\theta_{f1} = 0.636$, $\theta_{f2} = 0.3$, $\lambda_{fi} = \sqrt{2\kappa_{fi}\phi_{fi}}$, $\phi_{f1} = \phi_{f2} = 0.296$, and initial value $(f_{i0}, \sigma^2_{fi0}) = (0, \theta_{fi})$. Then we draw the factor loadings of efficient prices independently from a normal distribution with mean zero and unit variance.

The idiosyncratic components of efficient prices are generated as $U_{it} = \sigma_{it}\mathsf{W}^U_{it}$, where W^U_{it} is a Brownian motion and σ_{it} is generated by three different models for different i.

- For $1 \le i \le \lfloor d/3 \rfloor$, the volatility process is generated by an exponential ARCH diffusion limit model as in Nelson (1990):

$$\mathsf{d}\log(\sigma^2_{it}) = -0.6(0.157 - \log(\sigma^2_{it}))\mathsf{d}t + 0.25\mathsf{d}B^U_{it} \tag{4.12}$$

  with initial value $\log(\sigma^2_{i0}) = 0.157$, where B^U_{it} is a Brownian motion with $\mathsf{corr}(B^U_{it}, W^U_{it}) = -0.3$.

- For $\lfloor d/3 \rfloor + 1 \le i \le \lfloor 2d/3 \rfloor$, the volatility process is generated by a GARCH-M diffusion limit model as in Nelson (1990),

$$\mathsf{d}(\sigma^2_{it}) = (0.1 - \sigma^2_{it})\mathsf{d}t + 0.2\sigma^2_{it}\mathsf{d}B^U_{it} \tag{4.13}$$

  with initial value $\sigma^2_{i0} = 0.1$, where B^U_{it} is a Brownian motion with $\mathsf{corr}(B^U_{it}, W^U_{it}) = -0.3$.

- For $\lfloor 2d/3 \rfloor + 1 \le i \le d$, the volatility process is generated by a GARCH diffusion model as in Andersen and Bollerslev (1998),

$$\mathsf{d}(\sigma^2_{it}) = 0.035(0.636 - \sigma^2_{it})\mathsf{d}t + 0.2\sigma^2_{it}\mathsf{d}B^U_{it}, \tag{4.14}$$

  with initial value $\sigma^2_{i0} = 0.636$, where B^U_{it} is a Brownian motion with $\mathsf{corr}(B^U_{it}, W^U_{it}) = -0.3$.

The two dimensional Brownian motion (B^U_{it}, W^U_{it}) is independent over $1 \le i \le d$ and also independent with the driving Brownian motions (B_{1t}, W_{1t}) and (B_{2t}, W_{2t}) for factors of efficient prices.

As for microstructure noise, we introduce the noise-to-signal ξ_G^2 and ξ_V^2 as in Bollerslev et al. (2019), which take values $n^{2\bar{\tau}_G}$ and $n^{2\bar{\tau}_V}$, respectively, with $\bar{\tau}_G = -0.4, -0.6$ and $\bar{\tau}_V = -0.4, -0.6$. The variance of the factor for microstructure noise satisfies $\mathsf{Var}(G_t) = 0.5\xi_G^2(\frac{1}{nd}\sum_{i=1}^d\sum_{t=1}^n \sigma_{*,it}^4)^{1/2}$, and is thus time-invariant, where $\sigma_{*,it}$ is the spot volatility of the efficient price process of asset i at time t. The variance of idiosyncratic component V_{it} makes up $0.1\xi_V^2$ of the total variance, that is $\mathsf{Var}(V_{it}) = 0.1\xi_V^2(\frac{1}{n}\sum_{t=1}^n \sigma_{*,it}^4)^{1/2}$. We draw the factor G_t independently from a normal distribution with mean zero and variance $\mathsf{Var}(G_t)$, and draw V_{it} independently from a normal distribution with mean zero and variance $\mathsf{Var}(V_{it})$. Finally, we draw the factor loadings, λ_{Gi}, $i = 1, ..., d$, of microstructure noise independently from a normal distribution with mean one and unit variance.

We assume that the prices are synchronously recorded once every one or five minutes during 6.5 trading hours, that is $n = 390$ or 78. The number of assets is assumed to be $d = 50, 100, 300$ and 500. We present the simulation results based on 1000 Monte Carlo replications.

4.4. Simulation results

Firstly, we provide simulation results for the estimation of number of factors by the information criteria described in Section 4.1, as well as by the PANIC test with different significance levels, 1%, 5% and 10%. More specifically, we use Bai and Ng (2002)'s information criterion, IC_1, to estimate the total number of factors and then use the PANIC test to identify the number of factors for efficient prices. We compare this against Hallin and Liška (2007)'s information criteria $\mathsf{IC}_{2,i}$ and $\mathsf{IC}_{3,i}$, $i = 1, 2, 3$. For the PANIC test in (4.5), $\widehat{K}_H$ is determined by IC_1. We set $\overline{Q} = 10$ in both (4.9) and (4.10), and in Hallin and Liška (2007)'s information criteria, we set $B_n = 100$, $\iota = 0.5$, and $w(x) = 1 - |x|$ (i.e., the Bartlett weight function).

Tables 2 and 3 provide the average number of factors determined by each method (over 1000 replications) for $n = 78$ and $n = 390$, respectively. It can be seen that IC_1 has excellent performance in estimating the total number of factors in all scenarios. Other information criteria perform differently. Notice that in Section 4.1, $\mathsf{IC}_{2.i}$, $i = 1, 2, 3$, are used to estimate the total number of factors rather than factors for efficient prices. When $\bar{\tau}_V = -0.6$, i.e., when the idiosyncratic error of microstructure noise is relatively small, the average estimated number of factors by $\mathsf{IC}_{2.i}$, $i = 1, 2, 3$, is close to 2. When $\bar{\tau}_V$ increases to -0.4, this number decreases, especially for $\mathsf{IC}_{2.3}$ when $\bar{\tau}_G = -0.6$ (i.e., when the noise-to-signal ratio is relatively high). Indeed when $\bar{\tau}_G = -0.6$ and $\bar{\tau}_V = -0.4$, the average number of factors estimated from $\mathsf{IC}_{2.3}$ falls well below 1. The same pattern applies to $\mathsf{IC}_{3.i}$, $i = 1, 2, 3$. The results show that when used to estimate the total number of factors, the criteria $\mathsf{IC}_{2.i}$, $i = 1, 2, 3$, underestimate it in all scenarios. But similar to $\mathsf{IC}_{3.i}$, $i = 1, 2, 3$, they provide a good estimate of the number of factors for efficient prices when the idiosyncratic error of microstructure noise is relatively small. However, when the magnitude of the idiosyncratic error for microstructure noise increases, $\mathsf{IC}_{2.3}$ and $\mathsf{IC}_{3.3}$ seriously underestimate this number. The PANIC tests using the MQ statistic at different significance levels are denoted as $MQ_{1\%}$, $MQ_{5\%}$, and $MQ_{10\%}$ in Tables 2 and 3. They are used to determine the number of factors for efficient prices and perform satisfactorily in all scenarios, in particular for $MQ_{1\%}$. In summary, IC_1 is very satisfactory in determining the number of total factors, and the PANIC test with 1% significance level is the most robust method to determine the number of factors for efficient prices, outperforming $\mathsf{IC}_{2.2}$, which is the best performer among all of Hallin and Liška (2007)'s information criteria.

Next, we compare the estimation of common components in the dual factor model. Our method is denoted as DPCA*. For the alternative methods discussed in Section 4.2, we denote them as DPCA and PCA*-VECM, respectively. We use the relative estimation error (REE) to measure the performance of different methods. It is defined as

$$\text{REE} = \|\mathbf{M} - \widehat{\mathbf{M}}\|/\|\mathbf{M}\|,$$

where $\|\cdot\|$ can be the Frobenius norm, the max norm or the spectral norm, and $\widehat{\mathbf{M}}$ is an estimate of the matrix $\mathbf{M}$, which varies from place to place (depending on the quantity being estimated). In Tables 4–7, "diff.total" refers to the total common component, $\mathscr{h}\mathbf{\Lambda}_H^\intercal$. "EP" and "diff.EP" refer to the common components for efficient prices and efficient returns. That is, EP $= \mathscr{F}\mathbf{\Lambda}_F^\intercal$ and diff.EP $= \mathscr{f}\mathbf{\Lambda}_F^\intercal$, respectively. Similarly, "MN" and "diff.MN" refer to the common components for microstructure noise and its first difference, i.e., MN $= \mathscr{G}\mathbf{\Lambda}_G^\intercal$ and diff.MN $= \mathscr{g}\mathbf{\Lambda}_G^\intercal$, respectively. In the following study, we will only consider the cases $\bar{\tau}_G = \bar{\tau}_V = -0.6$ and $\bar{\tau}_G = \bar{\tau}_V = -0.4$. To focus on the comparison of different methods in estimating the common components, we assume that the true numbers of factors for efficient prices and microstructure noise are known. Otherwise, one can obtain correct estimates of the numbers of factors by IC_1 and the PANIC test in most cases (as can be seen from Tables 2 and 3).

Table 4 gives the simulation results when $n = 78$, $\bar{\tau}_G = -0.6$, $\bar{\tau}_V = -0.6$. In this case, the microstructure noise is smaller than the efficient returns in magnitude. We can see that the REE in all columns decreases when d increases. The estimates of the total common component, $\mathscr{h}\mathbf{\Lambda}_H^\intercal$, are the same for all the three methods, and the corresponding REE values are listed in the "diff.total" column. Since the estimates of $\mathscr{f}\mathbf{\Lambda}_F^\intercal$ and $\mathscr{g}\mathbf{\Lambda}_G^\intercal$ are decomposed from the estimate of $\mathscr{h}\mathbf{\Lambda}_H^\intercal$, the "diff.EP" and "diff.MN" have larger REEs than "diff.total" for all the three methods. Overall, DPCA* performs the best.

Table 5 presents the simulation results when $n = 78$, $\bar{\tau}_G = -0.4$, $\bar{\tau}_V = -0.4$. In this case, the microstructure noise is larger than the efficient returns in magnitude. Similar to Table 4, the REEs decrease when d increases. DPCA* performs the best, while DPCA performs the poorest. This may be due to the fact that DPCA* normalises the factors in the first step PCA, while DPCA does not. When the magnitude of the factors for microstructure noise, $\mathbf{D}_G\boldsymbol{g}_t$, is larger than that of the factors for efficient returns, $\boldsymbol{f}_t$, the magnitude of the cumulated factors for microstructure noise, $\mathbf{D}_G\boldsymbol{G}_t$, can still be larger than that of the factors for efficient prices, $\boldsymbol{F}_t$, and therefore, the factors corresponding to the leading eigenvalues in the second-step PCA may come from the microstructure noise.

Table 6 gives the simulation results when $n = 390$, $\bar{\tau}_G = -0.6$, and $\bar{\tau}_V = -0.6$. When we increase the sample size from 78 to 390, the REEs are smaller for "diff.total", "diff.EP", "diff.MN", and "EP".

Table 2: Average estimated number of factors and standard deviation (in parentheses) for sampling frequency=5 mins, true number of factors=2 (efficient prices)+1 (microstructure noise)

	IC_1	$\mathsf{IC}_{2.1}$	$\mathsf{IC}_{2.2}$	$\mathsf{IC}_{2.3}$	$\mathsf{IC}_{3.1}$	$\mathsf{IC}_{3.2}$	$\mathsf{IC}_{3.3}$	$MQ_{1\%}$	$MQ_{5\%}$	$MQ_{10\%}$
					$n = 78,\ \bar{\tau}_G = -0.4,\ \bar{\tau}_V = -0.4$					
d=50	3.026	1.910	1.976	1.419	1.874	1.984	1.477	1.998	1.931	1.855
	(0.025)	(0.082)	(0.023)	(0.260)	(0.110)	(0.064)	(0.284)	(0.030)	(0.086)	(0.150)
d=100	3.001	1.889	1.956	1.271	1.852	1.921	1.335	1.983	1.931	1.858
	(0.001)	(0.099)	(0.042)	(0.224)	(0.126)	(0.073)	(0.287)	(0.017)	(0.072)	(0.134)
d=300	3.000	1.889	1.951	1.183	1.846	1.924	1.195	1.986	1.942	1.871
	(0.000)	(0.099)	(0.047)	(0.176)	(0.130)	(0.070)	(0.239)	(0.014)	(0.059)	(0.122)
d=500	3.000	1.888	1.945	1.147	1.853	1.932	1.176	1.988	1.940	1.869
	(0.000)	(0.100)	(0.052)	(0.162)	(0.126)	(0.063)	(0.227)	(0.012)	(0.062)	(0.130)
					$n = 78,\ \bar{\tau}_G = -0.6,\ \bar{\tau}_V = -0.4$					
d=50	3.026	1.705	1.914	0.614	1.688	1.898	0.821	2.015	1.939	1.861
	(0.025)	(0.246)	(0.083)	(0.445)	(0.275)	(0.136)	(0.535)	(0.047)	(0.091)	(0.156)
d=100	3.001	1.647	1.847	0.277	1.626	1.797	0.492	1.984	1.930	1.858
	(0.001)	(0.287)	(0.140)	(0.255)	(0.312)	(0.182)	(0.400)	(0.018)	(0.073)	(0.134)
d=300	3.000	1.650	1.851	0.114	1.602	1.771	0.263	1.987	1.943	1.873
	(0.000)	(0.308)	(0.143)	(0.119)	(0.344)	(0.203)	(0.254)	(0.013)	(0.058)	(0.121)
d=500	3.000	1.652	1.834	0.083	1.590	1.788	0.180	1.988	1.939	1.869
	(0.000)	(0.337)	(0.171)	(0.086)	(0.408)	(0.227)	(0.178)	(0.012)	(0.063)	(0.130)
					$n = 78,\ \bar{\tau}_G = -0.4,\ \bar{\tau}_V = -0.6$					
d=50	3.001	2.000	2.000	1.983	1.997	2.020	1.955	1.991	1.949	1.884
	(0.001)	(0.000)	(0.000)	(0.017)	(0.005)	(0.024)	(0.043)	(0.011)	(0.056)	(0.117)
d=100	3.000	2.000	2.000	1.975	1.998	1.999	1.952	1.986	1.942	1.881
	(0.000)	(0.000)	(0.000)	(0.024)	(0.002)	(0.001)	(0.046)	(0.014)	(0.063)	(0.115)
d=300	3.000	2.000	2.000	1.990	2.000	2.000	1.969	1.991	1.951	1.890
	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.030)	(0.009)	(0.051)	(0.108)
d=500	3.000	2.000	2.000	1.990	2.000	2.000	1.974	1.988	1.948	1.887
	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.025)	(0.012)	(0.055)	(0.114)
					$n = 78,\ \bar{\tau}_G = -0.6,\ \bar{\tau}_V = -0.6$					
d=50	3.001	2.000	2.000	1.927	1.989	2.017	1.844	2.008	1.956	1.892
	(0.001)	(0.000)	(0.000)	(0.074)	(0.013)	(0.021)	(0.156)	(0.028)	(0.054)	(0.114)
d=100	3.000	1.999	2.000	1.915	1.992	1.996	1.847	1.987	1.945	1.881
	(0.000)	(0.001)	(0.000)	(0.086)	(0.008)	(0.004)	(0.152)	(0.015)	(0.060)	(0.115)
d=300	3.000	2.000	2.000	1.951	1.998	1.999	1.909	1.992	1.951	1.890
	(0.000)	(0.000)	(0.000)	(0.059)	(0.002)	(0.001)	(0.099)	(0.008)	(0.051)	(0.108)
d=500	3.000	2.000	2.000	1.951	2.000	2.000	1.914	1.988	1.949	1.887
	(0.000)	(0.000)	(0.000)	(0.057)	(0.000)	(0.000)	(0.109)	(0.012)	(0.054)	(0.114)

Table 3: Average estimated number of factors and standard deviation (in parentheses) for sampling frequency=1 min, true number of factors=2 (efficient prices)+1 (microstructure noise)

	IC_1	$\mathsf{IC}_{2.1}$	$\mathsf{IC}_{2.2}$	$\mathsf{IC}_{2.3}$	$\mathsf{IC}_{3.1}$	$\mathsf{IC}_{3.2}$	$\mathsf{IC}_{3.3}$	$MQ_{1\%}$	$MQ_{5\%}$	$MQ_{10\%}$
$n = 390$, $\bar{\tau}_G = -0.4$, $\bar{\tau}_V = -0.4$										
d=50	3.035	1.907	1.976	1.508	1.860	1.947	1.541	1.989	1.901	1.815
	(0.036)	(0.084)	(0.023)	(0.250)	(0.123)	(0.050)	(0.269)	(0.027)	(0.101)	(0.177)
d=100	3.004	1.913	1.964	1.266	1.843	1.913	1.388	1.987	1.917	1.827
	(0.004)	(0.080)	(0.035)	(0.195)	(0.132)	(0.080)	(0.264)	(0.015)	(0.080)	(0.165)
d=300	3.000	1.928	1.954	1.133	1.846	1.894	1.246	1.987	1.920	1.850
	(0.000)	(0.067)	(0.044)	(0.115)	(0.130)	(0.095)	(0.198)	(0.013)	(0.080)	(0.144)
d=500	3.000	1.924	1.967	1.076	1.846	1.896	1.176	1.988	1.923	1.850
	(0.000)	(0.070)	(0.032)	(0.070)	(0.130)	(0.093)	(0.155)	(0.012)	(0.071)	(0.146)
$n = 390$, $\bar{\tau}_G = -0.6$, $\bar{\tau}_V = -0.4$										
d=50	3.035	1.210	1.645	0.383	1.381	1.684	0.626	1.991	1.909	1.818
	(0.036)	(0.350)	(0.245)	(0.253)	(0.396)	(0.250)	(0.415)	(0.027)	(0.101)	(0.175)
d=100	3.004	1.034	1.395	0.081	1.201	1.453	0.317	1.990	1.920	1.827
	(0.004)	(0.281)	(0.291)	(0.075)	(0.419)	(0.358)	(0.253)	(0.010)	(0.078)	(0.165)
d=300	3.000	0.847	1.059	0.010	1.002	1.200	0.089	1.989	1.920	1.851
	(0.000)	(0.292)	(0.292)	(0.010)	(0.450)	(0.432)	(0.083)	(0.011)	(0.080)	(0.143)
d=500	3.000	0.777	0.955	0.004	0.906	1.049	0.057	1.988	1.923	1.850
	(0.000)	(0.306)	(0.319)	(0.004)	(0.440)	(0.443)	(0.054)	(0.012)	(0.071)	(0.146)
$n = 390$, $\bar{\tau}_G = -0.4$, $\bar{\tau}_V = -0.6$										
d=50	3.000	2.000	2.000	2.000	1.999	1.999	1.991	1.993	1.931	1.847
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.009)	(0.007)	(0.072)	(0.154)
d=100	3.000	2.000	2.000	2.000	2.000	2.000	1.999	1.991	1.932	1.848
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.009)	(0.067)	(0.147)
d=300	3.000	2.000	2.000	2.000	2.000	2.000	2.000	1.989	1.926	1.863
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.011)	(0.073)	(0.132)
d=500	3.000	2.000	2.000	2.000	2.000	2.000	2.000	1.990	1.931	1.860
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.010)	(0.064)	(0.137)
$n = 390$, $\bar{\tau}_G = -0.6$, $\bar{\tau}_V = -0.6$										
d=50	3.000	2.000	2.000	1.980	1.986	1.993	1.947	1.995	1.935	1.852
	(0.000)	(0.000)	(0.000)	(0.022)	(0.014)	(0.007)	(0.052)	(0.007)	(0.071)	(0.150)
d=100	3.000	2.000	2.000	1.996	1.998	1.999	1.961	1.991	1.935	1.849
	(0.000)	(0.000)	(0.000)	(0.004)	(0.002)	(0.001)	(0.038)	(0.009)	(0.065)	(0.146)
d=300	3.000	2.000	2.000	2.000	2.000	2.000	1.986	1.990	1.927	1.863
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.014)	(0.010)	(0.072)	(0.132)
d=500	3.000	2.000	2.000	2.000	2.000	2.000	1.995	1.989	1.931	1.861
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.011)	(0.064)	(0.136)

Table 4: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 78$, $\bar{\tau}_G = -0.6$, $\bar{\tau}_V = -0.6$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.209	0.261	0.283	0.284	0.690	0.266	0.339	0.359	0.731	0.306	0.319	0.278	1.006
	(0.021)	(0.111)	(0.059)	(0.092)	(0.383)	(0.116)	(0.106)	(0.134)	(0.371)	(0.080)	(0.121)	(0.113)	(0.381)
d=100	0.173	0.207	0.236	0.243	0.527	0.211	0.266	0.282	0.557	0.270	0.293	0.231	0.899
	(0.014)	(0.065)	(0.046)	(0.067)	(0.255)	(0.068)	(0.086)	(0.109)	(0.272)	(0.074)	(0.111)	(0.075)	(0.292)
d=300	0.143	0.166	0.199	0.209	0.389	0.168	0.202	0.208	0.403	0.245	0.273	0.200	0.859
	(0.010)	(0.038)	(0.041)	(0.061)	(0.173)	(0.042)	(0.075)	(0.091)	(0.216)	(0.084)	(0.117)	(0.066)	(0.354)
d=500	0.136	0.157	0.189	0.198	0.351	0.159	0.186	0.190	0.362	0.239	0.267	0.193	0.832
	(0.009)	(0.030)	(0.037)	(0.055)	(0.152)	(0.034)	(0.073)	(0.084)	(0.182)	(0.087)	(0.121)	(0.065)	(0.337)
REE under MAX norm													
d=50	0.210	0.253	0.271	0.284	0.583	0.259	0.338	0.363	0.620	0.293	0.317	0.266	0.854
	(0.049)	(0.100)	(0.082)	(0.113)	(0.276)	(0.103)	(0.142)	(0.154)	(0.276)	(0.103)	(0.141)	(0.104)	(0.356)
d=100	0.176	0.208	0.225	0.253	0.443	0.211	0.261	0.295	0.472	0.257	0.296	0.229	0.741
	(0.034)	(0.064)	(0.061)	(0.083)	(0.180)	(0.068)	(0.113)	(0.119)	(0.202)	(0.085)	(0.122)	(0.079)	(0.290)
d=300	0.158	0.176	0.196	0.240	0.332	0.177	0.200	0.235	0.338	0.236	0.290	0.203	0.684
	(0.029)	(0.043)	(0.054)	(0.078)	(0.119)	(0.046)	(0.081)	(0.099)	(0.162)	(0.088)	(0.122)	(0.070)	(0.325)
d=500	0.157	0.170	0.190	0.235	0.305	0.172	0.188	0.219	0.305	0.232	0.288	0.200	0.659
	(0.027)	(0.039)	(0.050)	(0.074)	(0.107)	(0.042)	(0.076)	(0.085)	(0.137)	(0.086)	(0.122)	(0.065)	(0.309)
REE under spectral norm													
d=50	0.161	0.257	0.235	0.230	0.660	0.261	0.315	0.302	0.677	0.259	0.258	0.274	0.983
	(0.022)	(0.126)	(0.067)	(0.079)	(0.389)	(0.130)	(0.131)	(0.122)	(0.375)	(0.087)	(0.099)	(0.129)	(0.379)
d=100	0.126	0.201	0.194	0.203	0.495	0.204	0.235	0.240	0.513	0.230	0.239	0.226	0.874
	(0.013)	(0.072)	(0.057)	(0.063)	(0.262)	(0.076)	(0.103)	(0.099)	(0.272)	(0.082)	(0.094)	(0.084)	(0.294)
d=300	0.115	0.161	0.166	0.188	0.356	0.163	0.167	0.184	0.368	0.215	0.228	0.197	0.834
	(0.010)	(0.041)	(0.053)	(0.060)	(0.176)	(0.045)	(0.082)	(0.078)	(0.212)	(0.088)	(0.095)	(0.074)	(0.353)
d=500	0.113	0.153	0.157	0.182	0.317	0.155	0.154	0.169	0.329	0.210	0.224	0.191	0.806
	(0.010)	(0.033)	(0.046)	(0.054)	(0.155)	(0.037)	(0.078)	(0.072)	(0.180)	(0.089)	(0.096)	(0.072)	(0.340)

Table 5: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 78$, $\bar{\tau}_G = -0.4$, $\bar{\tau}_V = -0.4$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.185	0.306	0.391	0.151	0.360	0.514	1.562	0.778	1.088	0.494	0.216	0.403	0.847
	(0.021)	(0.107)	(0.058)	(0.039)	(0.201)	(0.250)	(0.595)	(0.280)	(0.427)	(0.153)	(0.096)	(0.160)	(0.323)
d=100	0.151	0.476	1.478	0.742	1.049	0.257	0.322	0.131	0.295	0.434	0.199	0.364	0.803
	(0.014)	(0.235)	(0.652)	(0.313)	(0.484)	(0.069)	(0.050)	(0.031)	(0.140)	(0.144)	(0.085)	(0.155)	(0.282)
d=300	0.123	0.457	1.444	0.722	1.030	0.221	0.269	0.116	0.251	0.396	0.189	0.337	0.802
	(0.010)	(0.241)	(0.706)	(0.339)	(0.508)	(0.044)	(0.042)	(0.027)	(0.126)	(0.165)	(0.093)	(0.143)	(0.292)
d=500	0.117	0.214	0.256	0.112	0.246	0.450	1.435	0.717	1.042	0.387	0.186	0.333	0.789
	(0.010)	(0.041)	(0.043)	(0.026)	(0.134)	(0.241)	(0.723)	(0.346)	(0.524)	(0.179)	(0.098)	(0.158)	(0.322)
REE under MAX norm													
d=50	0.196	0.316	0.398	0.158	0.310	0.527	1.644	0.795	0.964	0.487	0.216	0.402	0.683
	(0.070)	(0.115)	(0.106)	(0.084)	(0.172)	(0.239)	(0.708)	(0.280)	(0.379)	(0.193)	(0.111)	(0.169)	(0.299)
d=100	0.157	0.489	1.548	0.759	0.929	0.273	0.322	0.139	0.247	0.422	0.196	0.367	0.636
	(0.041)	(0.231)	(0.753)	(0.303)	(0.431)	(0.080)	(0.080)	(0.043)	(0.102)	(0.167)	(0.087)	(0.159)	(0.267)
d=300	0.138	0.469	1.485	0.739	0.896	0.244	0.276	0.138	0.212	0.392	0.194	0.342	0.624
	(0.030)	(0.231)	(0.783)	(0.320)	(0.423)	(0.060)	(0.064)	(0.038)	(0.084)	(0.178)	(0.094)	(0.144)	(0.271)
d=500	0.134	0.241	0.269	0.138	0.208	0.465	1.487	0.734	0.909	0.384	0.192	0.338	0.616
	(0.028)	(0.059)	(0.062)	(0.038)	(0.087)	(0.234)	(0.785)	(0.325)	(0.448)	(0.179)	(0.096)	(0.150)	(0.298)
REE under operator norm													
d=50	0.119	0.297	0.317	0.122	0.346	0.485	1.775	0.739	0.950	0.444	0.184	0.400	0.839
	(0.020)	(0.119)	(0.068)	(0.036)	(0.203)	(0.276)	(0.711)	(0.271)	(0.384)	(0.191)	(0.090)	(0.179)	(0.323)
d=100	0.089	0.448	1.683	0.705	0.914	0.250	0.254	0.106	0.280	0.400	0.167	0.363	0.796
	(0.012)	(0.256)	(0.784)	(0.300)	(0.438)	(0.076)	(0.061)	(0.027)	(0.144)	(0.183)	(0.080)	(0.175)	(0.284)
d=300	0.078	0.432	1.644	0.687	0.898	0.215	0.219	0.099	0.235	0.386	0.159	0.341	0.795
	(0.008)	(0.264)	(0.842)	(0.324)	(0.455)	(0.049)	(0.052)	(0.023)	(0.130)	(0.205)	(0.088)	(0.162)	(0.291)
d=500	0.077	0.210	0.212	0.097	0.229	0.425	1.635	0.681	0.910	0.382	0.157	0.339	0.782
	(0.008)	(0.047)	(0.051)	(0.021)	(0.138)	(0.264)	(0.865)	(0.330)	(0.473)	(0.219)	(0.092)	(0.178)	(0.323)

Table 6: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 390$, $\bar{\tau}_G = -0.6$, $\bar{\tau}_V = -0.6$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.177	0.214	0.222	0.257	1.524	0.214	0.258	0.317	1.500	0.197	0.213	0.200	1.548
	(0.014)	(0.102)	(0.050)	(0.080)	(0.933)	(0.102)	(0.066)	(0.102)	(0.893)	(0.022)	(0.032)	(0.097)	(0.609)
d=100	0.131	0.156	0.166	0.193	1.117	0.156	0.184	0.223	1.113	0.153	0.170	0.150	1.238
	(0.008)	(0.069)	(0.034)	(0.058)	(0.663)	(0.069)	(0.043)	(0.069)	(0.654)	(0.019)	(0.031)	(0.067)	(0.425)
d=300	0.089	0.100	0.112	0.133	0.630	0.100	0.116	0.140	0.628	0.115	0.138	0.102	0.943
	(0.004)	(0.034)	(0.020)	(0.035)	(0.362)	(0.034)	(0.022)	(0.037)	(0.359)	(0.021)	(0.035)	(0.034)	(0.288)
d=500	0.078	0.087	0.098	0.117	0.494	0.087	0.097	0.114	0.491	0.106	0.131	0.092	0.863
	(0.003)	(0.026)	(0.018)	(0.030)	(0.285)	(0.026)	(0.016)	(0.026)	(0.284)	(0.021)	(0.035)	(0.026)	(0.253)
REE under MAX norm													
d=50	0.173	0.204	0.210	0.243	1.147	0.205	0.255	0.309	1.139	0.188	0.203	0.192	1.223
	(0.031)	(0.094)	(0.062)	(0.086)	(0.520)	(0.094)	(0.090)	(0.124)	(0.514)	(0.037)	(0.046)	(0.088)	(0.416)
d=100	0.125	0.147	0.154	0.182	0.833	0.148	0.178	0.215	0.831	0.141	0.162	0.142	0.947
	(0.020)	(0.060)	(0.041)	(0.063)	(0.375)	(0.060)	(0.058)	(0.081)	(0.368)	(0.030)	(0.041)	(0.058)	(0.302)
d=300	0.084	0.097	0.104	0.134	0.470	0.098	0.109	0.142	0.468	0.106	0.139	0.099	0.687
	(0.012)	(0.029)	(0.027)	(0.045)	(0.202)	(0.029)	(0.032)	(0.047)	(0.201)	(0.030)	(0.046)	(0.030)	(0.210)
d=500	0.077	0.087	0.092	0.125	0.369	0.087	0.090	0.121	0.366	0.100	0.136	0.091	0.629
	(0.011)	(0.025)	(0.024)	(0.041)	(0.155)	(0.025)	(0.022)	(0.033)	(0.152)	(0.028)	(0.043)	(0.025)	(0.201)
REE under operator norm													
d=50	0.151	0.214	0.191	0.214	1.516	0.214	0.244	0.262	1.477	0.170	0.183	0.201	1.543
	(0.016)	(0.116)	(0.050)	(0.057)	(0.934)	(0.116)	(0.080)	(0.091)	(0.900)	(0.026)	(0.022)	(0.113)	(0.609)
d=100	0.107	0.155	0.141	0.155	1.109	0.155	0.169	0.182	1.098	0.129	0.138	0.149	1.233
	(0.010)	(0.079)	(0.036)	(0.044)	(0.665)	(0.079)	(0.053)	(0.062)	(0.659)	(0.024)	(0.023)	(0.078)	(0.425)
d=300	0.062	0.098	0.093	0.107	0.620	0.098	0.099	0.114	0.617	0.095	0.111	0.100	0.937
	(0.004)	(0.040)	(0.028)	(0.034)	(0.365)	(0.040)	(0.031)	(0.036)	(0.362)	(0.028)	(0.034)	(0.039)	(0.289)
d=500	0.053	0.085	0.080	0.098	0.484	0.085	0.077	0.094	0.481	0.089	0.108	0.090	0.857
	(0.003)	(0.030)	(0.026)	(0.031)	(0.288)	(0.030)	(0.022)	(0.026)	(0.288)	(0.028)	(0.034)	(0.029)	(0.254)

Table 7: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 390$, $\bar{\tau}_G = -0.4$, $\bar{\tau}_V = -0.4$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.152	0.234	0.342	0.101	0.480	0.258	0.825	0.339	0.637	0.358	0.112	0.257	0.864
	(0.014)	(0.096)	(0.035)	(0.024)	(0.287)	(0.117)	(0.414)	(0.186)	(0.328)	(0.039)	(0.026)	(0.099)	(0.292)
d=100	0.111	0.177	0.250	0.074	0.353	0.196	0.593	0.242	0.477	0.278	0.090	0.211	0.805
	(0.007)	(0.064)	(0.021)	(0.012)	(0.204)	(0.088)	(0.357)	(0.160)	(0.273)	(0.040)	(0.022)	(0.081)	(0.273)
d=300	0.074	0.126	0.167	0.054	0.211	0.139	0.392	0.161	0.310	0.209	0.076	0.172	0.764
	(0.003)	(0.032)	(0.013)	(0.007)	(0.114)	(0.053)	(0.308)	(0.135)	(0.209)	(0.042)	(0.022)	(0.059)	(0.257)
d=500	0.064	0.115	0.145	0.048	0.176	0.128	0.332	0.136	0.264	0.192	0.073	0.163	0.752
	(0.003)	(0.024)	(0.012)	(0.006)	(0.095)	(0.053)	(0.302)	(0.132)	(0.211)	(0.045)	(0.022)	(0.055)	(0.266)
REE under MAX norm													
d=50	0.162	0.243	0.356	0.106	0.382	0.278	0.943	0.379	0.533	0.363	0.117	0.259	0.648
	(0.070)	(0.105)	(0.072)	(0.069)	(0.202)	(0.128)	(0.525)	(0.219)	(0.258)	(0.070)	(0.069)	(0.103)	(0.242)
d=100	0.110	0.184	0.253	0.073	0.269	0.209	0.670	0.270	0.386	0.269	0.088	0.212	0.574
	(0.034)	(0.063)	(0.047)	(0.033)	(0.133)	(0.086)	(0.436)	(0.185)	(0.223)	(0.057)	(0.037)	(0.079)	(0.216)
d=300	0.073	0.137	0.166	0.058	0.160	0.153	0.436	0.184	0.249	0.197	0.075	0.175	0.519
	(0.014)	(0.036)	(0.030)	(0.014)	(0.062)	(0.053)	(0.356)	(0.147)	(0.168)	(0.050)	(0.023)	(0.059)	(0.197)
d=500	0.066	0.131	0.147	0.057	0.132	0.144	0.365	0.158	0.212	0.183	0.073	0.168	0.514
	(0.013)	(0.035)	(0.028)	(0.014)	(0.051)	(0.053)	(0.333)	(0.140)	(0.167)	(0.049)	(0.022)	(0.056)	(0.208)
REE under operator norm													
d=50	0.104	0.231	0.291	0.091	0.476	0.250	0.925	0.321	0.572	0.299	0.102	0.254	0.860
	(0.015)	(0.110)	(0.037)	(0.021)	(0.286)	(0.128)	(0.502)	(0.188)	(0.309)	(0.039)	(0.022)	(0.113)	(0.291)
d=100	0.071	0.174	0.206	0.063	0.350	0.190	0.661	0.229	0.431	0.228	0.080	0.209	0.804
	(0.007)	(0.073)	(0.022)	(0.012)	(0.205)	(0.093)	(0.430)	(0.160)	(0.253)	(0.046)	(0.021)	(0.090)	(0.273)
d=300	0.041	0.122	0.127	0.041	0.207	0.134	0.430	0.153	0.280	0.178	0.064	0.171	0.763
	(0.003)	(0.036)	(0.017)	(0.007)	(0.115)	(0.054)	(0.369)	(0.134)	(0.191)	(0.058)	(0.022)	(0.067)	(0.257)
d=500	0.033	0.112	0.106	0.039	0.171	0.123	0.359	0.130	0.239	0.170	0.060	0.162	0.751
	(0.002)	(0.027)	(0.018)	(0.007)	(0.097)	(0.055)	(0.361)	(0.129)	(0.191)	(0.063)	(0.022)	(0.062)	(0.267)

However, all three methods have larger REEs for "MN". When $d = 50$, the REEs are even larger than 1, but they eventually decay when d becomes larger. Table 7 provides the simulation results when $n = 390$, $\bar{\tau}_G = -0.4$, and $\bar{\tau}_V = -0.4$. The pattern is similar to that in Table 5, i.e., DPCA* outperforms DPCA and PCA*-VECM.

5. Application

We now apply the proposed method to 5-min intraday returns of S&P 500 Index constituents (505 stocks in total). The data were collected from Bloomberg from 11th January, 2021 to 14th May, 2021. For each day, the observed prices constitute an $(n+1)$-by-d matrix, $\mathscr{X}$, with $n = 78$ and $d \leq 505$. The value of d (i.e., the number of stocks) may vary from day to day since we discard less active stocks which did not trade in any one of the 5-min observation intervals.

Firstly, we determine the number of factors using IC_1 and the PANIC test for each of the 87 trading days within the sampling period. Figure 1 shows that the PANIC tests with different significance levels give different numbers of factors for efficient prices. The estimated total number of factors is 6.253 ± 1.193 (mean $\pm$ standard deviation) on average. The estimated number of factors for microstructure noise is 0.805 ± 0.975 from 1% PANIC tests, 1.494 ± 1.209 for 5%, and 1.908 ± 1.300 for 10% tests. Nearly half of the days (40 in 87) have non-zero common components for microstructure noise, and 26 have two or more factors. On the days of 4th February, 28th February, 29th April, and 11th May, the number of factors for microstructure noise is the highest at 3.

To give a specific example, let us consider the results for 26th April 2021. The estimated total number of factors is 6, among which, 4 are identified as factors for efficient prices by the PANIC test at 1% significance level, 3 at 5% significance level, and 1 at 10% significance level. Figure 2 shows the 6 estimated factors in cumulative form, i.e., $\left(\widehat{\boldsymbol{\beta}}_{\perp}, \widehat{\boldsymbol{\beta}}\right)^{\intercal} \widehat{H}^{*}_{s\Delta}$, where $\left(\widehat{\boldsymbol{\beta}}_{\perp}, \widehat{\boldsymbol{\beta}}\right)$ is the matrix of eigenvectors of the matrix $n^{-1}\widehat{\mathscr{H}}^{*c\intercal}\widehat{\mathscr{H}}^{*c}$, arranged in descending order of their corresponding eigenvalues. The first factor appears most likely to be nonstationary. The next two or three factors may be nonstationary too. The last two factors are the most stable among the six.

Lastly, we look at the decomposition of prices (aggregated returns) into three components: the common component of efficient prices, $\boldsymbol{\Lambda}_F \mathbf{F}_t$ (CC.EP), the common component of microstructure noise, $\boldsymbol{\Lambda}_G \mathbf{D}_G \mathbf{G}_t$ (CC.MN), and idiosyncratic errors (residuals). We show this through the first five of the alphabetically ordered 505 stocks that have the stock ticker symbols – A, AAL, AAP, AAPL, and ABBV. Figures 3 and 4 give the decompositions of the prices of the five stocks on 26th April 2021. Figure 3 shows the decomposition when we choose $\widehat{K}_F = 4$ and $\widehat{K}_G = 2$, and Figure 4 shows the decomposition when $\widehat{K}_F = 3$ and $\widehat{K}_G = 3$ (chosen by PANIC test at 5% significance level). The common component of the microstructure noise can only explain a small amount of the variability of the prices. When $\widehat{K}_G = 1$ the common components of the microstructure noise for most stocks are negligible. When $\widehat{K}_G = 3$, the common components of the microstructure noise are more variable, but less variable than the residuals. In summary, our analysis finds existence of common components for the microstructure noise of S&P 500 stocks, although their magnitude is small. The small magnitude is also consistent with the expectation that there are very few arbitrage opportunities in a frictional market.

6. Conclusion and Future work

We consider a dual-factor model for high-frequency stock prices contaminated with microstructure noise. We develop the Double Principle Component Analysis (DPCA*) to estimate the separate factor structures for efficient prices and microstructure noise. When comparing with the PCA-VECM approach, the DPCA* approach is free from the need to impose strong parametric assumptions on the microstructure noise and applies instead to a broad class of stationary processes. The estimators are proven to be consistent and perform well in simulations. The empirical analysis of intraday returns of S&P 500 constituents provides some evidence of co-movement in the microstructure noise, apart from co-movement of prices caused by common systematic risk factors.

Identifying and separating out common components of microstructure noise from the prices are very useful for the study of properties of the microstructure noise and efficient price processes. For example, once the common components for microstructure noise are separated out, the common components for efficient prices are no longer contaminated by microstructure noise and hence, can be used to obtain a more accurate estimate of the common part of realized volatility. For the idiosyncratic part of realized volatility, one can use the estimated idiosyncratic errors and apply the pre-averaging method of Jacod et al. (2009). Adding these two parts together, we get an estimator of the realized volatility matrix. We may introduce sparsity or block structure into idiosyncratic components like Dai et al. (2019) and Aït-Sahalia and Xiu (2017), respectively. However, since our main interests are the identification of common factors, we avoid introducing these structures and leave the estimation of the realized volatility matrix to the future work.

The estimated common factors and loadings for microstructure noise provide useful tools for portfolio management. With such estimates, portfolio managers can construct a new factor mimicking portfolio which is only exposed to the factors of microstructure noise. Such a portfolio can be used to hedge risks from microstructure noise. Since the portfolio return is usually stationary, one can apply the mean-reverting strategy to earn profits from the portfolio, once its volatility is large enough to cover the cost. Even if its volatility is small, one can still time the market according to it, e.g., when adjusting the position of a portfolio, to lower the cost.

References

Aït-Sahalia, Y., Kalnina, I., Xiu, D., 2020. High-frequency factor models and regressions. Journal of Econometrics.

Aït-Sahalia, Y., Xiu, D., 2017. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. Journal of Econometrics 201 (2), 384–399.

Andersen, T. G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 885–905.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71 (1), 135–171.

Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. The Annals of Statistics 40 (1), 436–465.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.

Bai, J., Ng, S., 2004. A panic attack on unit roots and cointegration. Econometrica 72 (4), 1127–1177.

Barigozzi, M., Hallin, M., 2016. Generalized dynamic factor models and volatilities: recovering the market volatility shocks. The Econometrics Journal 19 (1), 33–60.

Barigozzi, M., Hallin, M., 2017. Generalized dynamic factor models and volatilities: estimation and forecasting. Journal of Econometrics 201 (2), 307–321.

Barigozzi, M., Lippi, M., Luciani, M., 2020a. Cointegration and error correction mechanisms for singular stochastic vectors. Econometrics 8 (1), 3.

Barigozzi, M., Lippi, M., Luciani, M., 2020b. Large-dimensional dynamic factor models: Estimation of impulse–response functions with I (1) cointegrated factors. Journal of Econometrics.

Bollerslev, T., Meddahi, N., Nyawa, S., 2019. High-dimensional multivariate realized volatility estimation. Journal of Econometrics 212 (1), 116–136.

Briggs, N. E., MacCallum, R. C., 2003. Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. Multivariate Behavioral Research 38 (1), 25–56.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52 (1), 57–82.

Choi, J. Y., Salandro, D., Shastri, K., 1988. On the estimation of bid-ask spreads: Theory and evidence. Journal of Financial and Quantitative Analysis 23 (2), 219–230.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. Journal of Financial Economics 56 (1), 3–28.

Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. Journal of Econometrics 159 (1), 116–133.

Dai, C., Lu, K., Xiu, D., 2019. Knowing factors or factor loadings, or neither? Evaluating estimators of large covariance matrices with noisy and asynchronous data. Journal of Econometrics 208 (1), 43–79.

Davis, C., Kahan, W. M., 1970. The rotation of eigenvectors by a perturbation. iii. SIAM Journal on Numerical Analysis 7 (1), 1–46.

Delattre, S., Jacod, J., 1997. A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. Bernoulli, 1–28.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116 (1), 1–22.

Fan, J., Furger, A., Xiu, D., 2016. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. Journal of Business & Economic Statistics 34 (4), 489–503.

Fan, J., Kim, D., 2018. Robust high-dimensional volatility matrix estimation for high-frequency factor model. Journal of the American Statistical Association 113 (523), 1268–1283.

Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. Journal of the American Statistical Association 107 (497), 412–428.

Fan, J., Liao, Y., Mincheva, M., 2011. High dimensional covariance matrix estimation in approximate factor models. Annals of statistics 39 (6), 3320.

Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75 (4), 603–680.

Glosten, L. R., 1987. Components of the bid-ask spread and the statistical properties of transaction prices. The Journal of Finance 42 (5), 1293–1307.

Glosten, L. R., Harris, L. E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21 (1), 123–142.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102 (478), 603–617.

Hallin, M., Liška, R., 2011. Dynamic factors in the presence of blocks. Journal of Econometrics 163 (1), 29–41.

Harris, D., 1997. Principal components analysis of cointegrated time series. Econometric Theory 13 (4), 529–557.

Hasbrouck, J., 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. The Review of Financial Studies 6 (1), 191–212.

Hasbrouck, J., Ho, T. S., 1987. Order arrival, quote behavior, and the return-generating process. The Journal of Finance 42 (4), 1035–1048.

Hasbrouck, J., Seppi, D. J., 2001. Common factors in prices, order flows, and liquidity. Journal of Financial Economics 59 (3), 383–411.

Herwartz, H., Lütkepohl, H., 2011. Generalized least squares estimation for cointegration parameters under conditional heteroskedasticity. Journal of Time Series Analysis 32 (3), 281–291.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Processes and Their Applications 119 (7), 2249–2276.

Johansen, S., 1995. Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press on Demand.

Kim, D., Wang, Y., Zou, J., 2016. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. Stochastic Processes and their Applications 126 (11), 3527–3577.

Kunitomo, N., Kurisu, D., 2019. Detecting factors of quadratic variation in the presence of market microstructure noise.

Kunitomo, N., Sato, S., 2013. Separating information maximum likelihood estimation of the integrated volatility and covariance with micro-market noise. The North American Journal of Economics and Finance 26, 282–309.

Li, Y., Mykland, P. A., 2007. Are volatility estimators robust with respect to modeling assumptions? Bernoulli 13 (3), 601–622.

Merlevède, F., Peligrad, M., Rio, E., 2011. A bernstein type inequality and moderate deviations for weakly dependent sequences. Probability Theory and Related Fields 151 (3-4), 435–474.

Nelson, D. B., 1990. ARCH models as diffusion approximations. Journal of Econometrics 45 (1-2), 7–38.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. The Review of Economics and Statistics 92 (4), 1004–1016.

Pelger, M., 2019. Large-dimensional factor modeling based on high-frequency observations. Journal of Econometrics 208 (1), 23–42.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. The Journal of Finance 39 (4), 1127–1139.

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13 (3), 341–360.

Seo, B., 2007. Asymptotic distribution of the cointegrating vector estimator in error correction models with conditional heteroskedasticity. Journal of Econometrics 137 (1), 68–111.

Sharpe, W. F., 1994. The sharpe ratio. Journal of portfolio management 21 (1), 49–58.

Stock, J. H., Watson, M. W., 1988. Testing for common trends. Journal of the American statistical Association 83 (404), 1097–1107.

Tao, M., Wang, Y., Chen, X., 2013a. Fast convergence rates in estimating large volatility matrices using high-frequency financial data. Econometric Theory 29 (4), 838–856.

Tao, M., Wang, Y., Yao, Q., Zou, J., 2011. Large volatility matrix inference via combining low-frequency and high-frequency approaches. Journal of the American Statistical Association 106 (495), 1025–1040.

Tao, M., Wang, Y., Zhou, H. H., 2013b. Optimal sparse volatility matrix estimation for high-dimensional itô processes with measurement errors. The Annals of Statistics 41 (4), 1816–1864.

Wang, Y., Zou, J., 2010. Vast volatility matrix estimation for high-frequency financial data. The Annals of Statistics 38 (2), 943–978.

Yu, Y., Wang, T., Samworth, R. J., 2015. A useful variant of the davis–kahan theorem for statisticians. Biometrika 102 (2), 315–323.

Appendix A: Technical Proofs

In the subsequent proofs, we often make use of the following Weyl's inequality, for two $n \times n$ symmetric matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, with eigenvalues $\mu_j(\mathbf{M}_1)$ and $\mu_j(\mathbf{M}_2)$:

$$|\mu_j(\mathbf{M}_1) - \mu_j(\mathbf{M}_2)| \leq \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{sp}}, \tag{A.1}$$

for $j = 1, ..., n$. If $\mathbf{M}_1$ and $\mathbf{M}_2$ are invertible and $\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{sp}}\|\mathbf{M}_2^{-1}\|_{\mathsf{sp}} < 1$, we have

$$\begin{aligned}
\|\mathbf{M}_1^{-1} - \mathbf{M}_2^{-1}\|_{\mathsf{sp}} &\leq \|\mathbf{M}_1^{-1}\|_{\mathsf{sp}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{sp}}\|\mathbf{M}_2^{-1}\|_{\mathsf{sp}} \\
&\leq \|\mathbf{M}_1^{-1} - \mathbf{M}_2^{-1}\|_{\mathsf{sp}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{sp}}\|\mathbf{M}_2^{-1}\|_{\mathsf{sp}} + \|\mathbf{M}_2^{-1}\|_{\mathsf{sp}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{sp}}\|\mathbf{M}_2^{-1}\|_{\mathsf{sp}} \\
&\leq \frac{\|\mathbf{M}_2^{-1}\|_{\mathsf{sp}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{sp}}\|\mathbf{M}_2^{-1}\|_{\mathsf{sp}}}{1 - \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{sp}}\|\mathbf{M}_2^{-1}\|_{\mathsf{sp}}}.
\end{aligned} \tag{A.2}$$

Note that the max norm is not submultiplicative, but we can use $\|\mathbf{M}_1\mathbf{M}_2\|_{\mathsf{MAX}} \leq \|\mathbf{M}_1\|_\infty\|\mathbf{M}_2\|_{\mathsf{MAX}}$ or $\|\mathbf{M}_1\mathbf{M}_2\|_{\mathsf{MAX}} \leq \|\mathbf{M}_1\|_{\mathsf{MAX}}\|\mathbf{M}_2\|_1$.

Lemma A.1. *Under Assumptions 1–4, 5*, 6 and 7, we have, for each $1 \leq j \leq K_H$,*
(i)

$$\left|\xi_j - \frac{\boldsymbol{b}_j}{\|\boldsymbol{b}_j\|_2}\right| = O\left(\frac{m_{U,d} + n^{2\bar{\tau}_V^*} m_{v,d}}{dn^{2\underline{\tau}_G^{*-}}}\right), \tag{A.3}$$

where ξ_j is the eigenvector of $\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^\intercal$ corresponding to the jth largest eigenvalue.
(ii) If, in addition, Assumption 8(i) holds, then $\mu_{K_H}(\boldsymbol{x}^\intercal\boldsymbol{x}) \geq Cdn^{2\underline{\tau}_G^{*-}}$.

Proof. Recall that

$$\mathbf{\Sigma}_x = \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^\intercal + \mathbf{\Sigma}_w.$$

Let $\mathbf{B} = \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}\mathbf{Q} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{K_H})$ with $\|\boldsymbol{b}_j\|_2$'s sorted in a descending order, where $\mathbf{Q}$ is an orthogonal matrix such that $\mathbf{Q}^\intercal\mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}\mathbf{Q}$ is a diagonal matrix. Then $\|\boldsymbol{b}_j\|_2^2$, $1 \leq j \leq K_H$, are the non-zero eigenvalues of $\mathbf{B}\mathbf{B}^\intercal = \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^\intercal$ and also the eigenvalues of $\mathbf{B}^\intercal\mathbf{B} = \mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}$. Therefore,

$$\|\boldsymbol{b}_j\|_2^2 \leq \|\boldsymbol{b}_1\|_2^2 = \|\mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}\|_{\mathsf{sp}} \leq \|\mathbf{D}_H\|_{\mathsf{sp}}^2\|\mathbf{\Sigma}_h\|_{\mathsf{sp}} \cdot \|\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\|_{\mathsf{sp}} = O(dn^{2\bar{\tau}_G^{*+}}),$$

where the last equality holds by Assumptions 2 and 4. On the other hand,

$$\begin{aligned}
\|\boldsymbol{b}_{K_H}\|_2^2 &= \mu_{K_H}(\mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}) \\
&\geq \mu_{K_H}(\mathbf{\Sigma}_h)\mu_{K_H}^2(\mathbf{D}_H)\mu_{K_H}(\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H) \geq Cdn^{2\underline{\tau}_G^{*-}}.
\end{aligned} \tag{A.4}$$

By the Sin theta theorem in Davis and Kahan (1970) (see Yu et al. (2015) for a statistician-friendly version), we have

$$\left|\xi_j - \frac{\boldsymbol{b}_j}{\|\boldsymbol{b}_j\|_2}\right| \leq \frac{\sqrt{2}\|\boldsymbol{\Sigma}_w\|_{\mathsf{sp}}}{\min\left\{\left|\mu_{j-1}(\boldsymbol{\Sigma}_x) - \|\boldsymbol{b}_j\|_2^2\right|, \left|\mu_{j+1}(\boldsymbol{\Sigma}_x) - \|\boldsymbol{b}_j\|_2^2\right|\right\}}, \tag{A.5}$$

with the convention $\mu_0(\cdot) = \infty$. By Weyl's inequality and triangle inequality, we have

$$|\mu_j(\boldsymbol{\Sigma}_x) - \|\boldsymbol{b}_j\|_2^2| \leq \|\boldsymbol{\Sigma}_w\|_{\mathsf{sp}} = O(m_{w,nd}), \tag{A.6}$$

for $1 \leq j \leq K_H$ and

$$0 < \mu_j(\boldsymbol{\Sigma}_x) \leq \|\boldsymbol{\Sigma}_w\|_{\mathsf{sp}} = O(m_{w,nd}),$$

for $K_H + 1 \leq j \leq d$, where $m_{w,nd} = m_{U,d} + n^{2\bar{\tau}_V^*} m_{v,d}$. Using the triangle inequality, (A.6) and Assumption 7, we have

$$\begin{aligned}\left|\mu_{j-1}(\boldsymbol{\Sigma}_x) - \|\boldsymbol{b}_j\|_2^2\right| &\geq \left|\|\boldsymbol{b}_{j-1}\|_2^2 - \|\boldsymbol{b}_j\|_2^2\right| - \left|\mu_{j-1}(\boldsymbol{\Sigma}_x) - \|\boldsymbol{b}_{j-1}\|_2^2\right| \\ &\geq Cdn^{2\underline{\tau}_G^{*-}},\end{aligned}$$

for $1 \leq j \leq K_H$, as $m_{w,nd} = o(dn^{2\underline{\tau}_G^{*-}})$ under Assumption 5*. Similarly, $|\mu_{j+1}(\boldsymbol{\Sigma}_x) - \|\boldsymbol{b}_j\|_2^2| \geq Cdn^{2\underline{\tau}_G^{*-}}$, when $1 \leq j \leq K_H - 1$. Therefore, by (A.5), we can prove (i).For part (ii), by Weyl's inequality,

$$|\mu_{K_H}(\boldsymbol{x}^\intercal \boldsymbol{x}) - \mu_{K_H}(\boldsymbol{\Sigma}_x)| \leq \|\boldsymbol{x}^\intercal \boldsymbol{x} - \boldsymbol{\Sigma}_x\|_{\mathsf{sp}}.$$

Then using $\mu_{K_H}(\boldsymbol{\Sigma}_x) \geq \|\boldsymbol{b}_{K_H}\|_2^2$, we have

$$\begin{aligned}\mu_{K_H}(\boldsymbol{x}^\intercal \boldsymbol{x}) &\geq \mu_{K_H}(\boldsymbol{\Sigma}_x) - \|\boldsymbol{x}^\intercal \boldsymbol{x} - \boldsymbol{\Sigma}_x\|_{\mathsf{sp}} \\ &\geq \|\boldsymbol{b}_{K_H}\|_2^2 - \|\boldsymbol{x}^\intercal \boldsymbol{x} - \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal\|_{\mathsf{sp}} - \|\boldsymbol{\Sigma}_w\|_{\mathsf{sp}}, \\ &\quad -\|\boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal - \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\Sigma}_h \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal\|_{\mathsf{sp}}\end{aligned}$$

where $\boldsymbol{\Sigma}_w = \boldsymbol{\Sigma}_U + n\mathbf{D}_V \boldsymbol{\Sigma}_v \mathbf{D}_V$. Therefore by (A.4), (A.6) and Assumption 5*, we only need to show that

$$\|\boldsymbol{x}^\intercal \boldsymbol{x} - \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal\|_{\mathsf{sp}} = o_P(dn^{2\underline{\tau}_G^{*-}}), \tag{A.7}$$

and

$$\|\boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal - \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\Sigma}_h \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal\|_{\mathsf{sp}} = o_P(dn^{2\underline{\tau}_G^{*-}}). \tag{A.8}$$

As for (A.7), using Lemma B.4, we have

$$\begin{aligned}\|\boldsymbol{x}^\intercal \boldsymbol{x} - \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal\|_{\mathsf{sp}} &\leq 2\|\boldsymbol{\Lambda}_H\|_{\mathsf{sp}} \|\mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{w}\|_{\mathsf{sp}} + \|\boldsymbol{w}^\intercal \boldsymbol{w} - \boldsymbol{\Sigma}_w\|_{\mathsf{sp}} + \|\boldsymbol{\Sigma}_w\|_{\mathsf{sp}} \\ &= O_P(d(\log d/n)^{1/2} \cdot n^{\bar{\tau}_V^{*+} + \bar{\tau}_G^{*+}}) \\ &\quad + O_P(d(\log d/n)^{1/2} \cdot n^{2\bar{\tau}_V^{*+}}) + O(m_{w,nd}),\end{aligned} \tag{A.9}$$

since $\|\mathbf{\Lambda}_H\|_{\mathsf{sp}} = O(d^{1/2})$, $\|\mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{w}\|_{\mathsf{sp}} \leq d^{1/2}\|\mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{w}\|_1$ and $\|\boldsymbol{w}^\intercal \boldsymbol{w} - \mathbf{\Sigma}_w\|_{\mathsf{sp}} \leq d\|\boldsymbol{w}^\intercal \boldsymbol{w} - \mathbf{\Sigma}_w\|_{\mathsf{MAX}}$. Then, under Assumptions 5* and 8(i),

$$\|\boldsymbol{x}^\intercal \boldsymbol{x} - \mathbf{\Lambda}_H \mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \mathbf{\Lambda}_H^\intercal\|_{\mathsf{sp}} = o_P(dn^{2\underline{\tau}_G^{*-}}),$$

if $n^{1+4\underline{\tau}_G^{*-}-2\bar{\tau}_V^{*+}-2(\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+})}/(\log d) \to \infty$ and $m_{w,nd} = o(dn^{2\underline{\tau}_G^{*-}})$. As for (A.8), by Assumptions 4 and 8(i), we have

$$\begin{aligned}
& \|\mathbf{\Lambda}_H \mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \mathbf{\Lambda}_H^\intercal - \mathbf{\Lambda}_H \mathbf{D}_H \mathbf{\Sigma}_h \mathbf{D}_H \mathbf{\Lambda}_H^\intercal\|_{\mathsf{sp}} \\
\leq\ & C_\Lambda d \|\mathbf{D}_H \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H - \mathbf{D}_H \mathbf{\Sigma}_h \mathbf{D}_H\|_{\mathsf{sp}} \\
\leq\ & K_H C_\Lambda d \cdot \max\left\{\|\boldsymbol{f}^\intercal \boldsymbol{f} - \mathbf{\Sigma}_F\|_{\mathsf{MAX}}, \|\mathbf{D}_G \boldsymbol{g}^\intercal \boldsymbol{g} \mathbf{D}_G - \mathbf{D}_G \mathbf{\Sigma}_g \mathbf{D}_G\|_{\mathsf{MAX}}, \|\boldsymbol{f}^\intercal \boldsymbol{g} \mathbf{D}_G\|_{\mathsf{MAX}}\right\} \\
=\ & O_P(dn^{2\bar{\tau}_G^{*+}}(\log d/n)^{1/2}) = o_P(dn^{2\underline{\tau}_G^{*-}}).
\end{aligned}$$

where the last line follows from Lemmas B.1(ii), B.2(ii) and B.3(iv). Hence we complete the proof. □

Lemma A.2. *Suppose that Assumptions 1–8 are satisfied. We have*
(i)

$$\left\|\widehat{\mathbf{\Lambda}}_H - \mathbf{\Lambda}_H \mathbf{D}_H \mathbf{R}\right\|_{\mathsf{MAX}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right) = o_P(1), \tag{A.10}$$

where

$$a_{nd} = (\log d)^{1/2} \frac{n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+}}}{n^{1/2}} + \frac{m_{w,nd}}{d^{1/2}},$$

in which $\bar{\tau}_G^{*+} = (1/2+\bar{\tau}_G)_+$, $\underline{\tau}_G^{*-} = (1/2+\underline{\tau}_G)_-$, $\bar{\tau}_V^{*+} = (1/2+\bar{\tau}_V)_+$, $m_{w,nd} = m_{U,d} + n^{2\bar{\tau}_V^*} m_{v,d}$, and $\bar{\tau}_V^* = (1/2+\bar{\tau}_V)$.
(ii)

$$\|\mathbf{D}_H \mathbf{R}\|_{\mathsf{sp}} = O_P(1), \qquad \|(\mathbf{D}_H \mathbf{R})^{-1}\|_{\mathsf{sp}} = O_P(1), \tag{A.11}$$

and

$$d^{-1/2}\|\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\|_{\mathsf{sp}} = O_P(1). \tag{A.12}$$

(iii)

$$\left\|\widehat{\boldsymbol{\hbar}}^\intercal - (\mathbf{D}_H \mathbf{R})^{-1} \mathbf{D}_H \boldsymbol{\hbar}^\intercal\right\|_{\mathsf{MAX}} \leq \left\|\widehat{\boldsymbol{\hbar}}^\intercal - (\mathbf{D}_H \mathbf{R})^{-1} \mathbf{D}_H \boldsymbol{\hbar}^\intercal\right\|_{\mathsf{sp}} = O_P\left(n^{-\underline{\tau}_G^{*-}} \cdot a_{nd}\right), \tag{A.13}$$

where the rotation matrix $\mathbf{R}$ is defined as

$$\mathbf{R} = \boldsymbol{\hbar}^\intercal \boldsymbol{\hbar} \mathbf{D}_H \mathbf{\Lambda}_H^\intercal \widehat{\mathbf{\Lambda}}_H \widehat{\mathbf{D}}_{x,K_H}^{-1}, \tag{A.14}$$

in which $\widehat{\mathbf{D}}_{x,K_H}$ is a $K_H \times K_H$ diagonal matrix with the diagonal elements being the first K_H largest eigenvalues of $\boldsymbol{x}^\intercal \boldsymbol{x}$ arranged in descending order.

Proof of Lemma A.2. By the definition of PCA estimation, we may show that

$$\begin{aligned}
\left(\widehat{\boldsymbol{\Lambda}}_H - \boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R}\right)\widehat{\mathbf{D}}_{x,K_H} &= (\boldsymbol{x}^\intercal\boldsymbol{x} - \boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\hbar}^\intercal\boldsymbol{\hbar}\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal)\widehat{\boldsymbol{\Lambda}}_H \\
&= \boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\hbar}^\intercal\boldsymbol{w}\widehat{\boldsymbol{\Lambda}}_H + \boldsymbol{w}^\intercal\boldsymbol{\hbar}\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal\widehat{\boldsymbol{\Lambda}}_H \\
&\quad + (\boldsymbol{w}^\intercal\boldsymbol{w} - \boldsymbol{\Sigma}_w)\widehat{\boldsymbol{\Lambda}}_H + \boldsymbol{\Sigma}_w\widehat{\boldsymbol{\Lambda}}_H.
\end{aligned}$$

For the first term on the right hand side of the second equality, we have, by similar argument to (A.9),

$$\begin{aligned}
\|\boldsymbol{x}^\intercal\boldsymbol{x} - \boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\hbar}^\intercal\boldsymbol{\hbar}\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal\|_{\mathsf{MAX}} &\le 2\|\boldsymbol{\Lambda}_H\|_{\mathsf{MAX}}\|\mathbf{D}_H\boldsymbol{\hbar}^\intercal\boldsymbol{w}\|_1 + \|\boldsymbol{w}^\intercal\boldsymbol{w} - \boldsymbol{\Sigma}_w\|_{\mathsf{MAX}} + \|\boldsymbol{\Sigma}_w\|_{\mathsf{MAX}} \\
&= O_P((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}}) \\
&\quad + O_P((\log d/n)^{1/2}\cdot n^{2\bar{\tau}_V^{*+}}) + O(m_{w,nd}).
\end{aligned}$$

For the remaining terms, we have

$$\begin{aligned}
&\|\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\hbar}^\intercal\boldsymbol{w}\widehat{\boldsymbol{\Lambda}}_H\|_{\mathsf{MAX}} \le \|\boldsymbol{\Lambda}_H\|_{\mathsf{MAX}}\|\mathbf{D}_H\boldsymbol{\hbar}^\intercal\boldsymbol{w}\|_1\|\widehat{\boldsymbol{\Lambda}}_H\|_1 = O_P(d(\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}}), \\
&\|\boldsymbol{w}^\intercal\boldsymbol{\hbar}\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal\widehat{\boldsymbol{\Lambda}}_H\|_{\mathsf{MAX}} \le \|\boldsymbol{w}^\intercal\boldsymbol{\hbar}\mathbf{D}_H\|_{\mathsf{MAX}}\|\boldsymbol{\Lambda}_H^\intercal\|_1\|\widehat{\boldsymbol{\Lambda}}_H\|_1 = O_P(d(\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}}), \\
&\|(\boldsymbol{w}^\intercal\boldsymbol{w} - \boldsymbol{\Sigma}_w)\widehat{\boldsymbol{\Lambda}}_H\|_{\mathsf{MAX}} \le \|\boldsymbol{w}^\intercal\boldsymbol{w} - \boldsymbol{\Sigma}_w\|_{\mathsf{MAX}}\|\widehat{\boldsymbol{\Lambda}}_H\|_1 = O_P(d(\log d/n)^{1/2}\cdot n^{2\bar{\tau}_V^{*+}}),
\end{aligned}$$

and

$$\|\boldsymbol{\Sigma}_w\widehat{\boldsymbol{\Lambda}}_H\|_{\mathsf{MAX}} \le \|\boldsymbol{\Sigma}_w\|_\infty\|\widehat{\boldsymbol{\Lambda}}_H\|_{\mathsf{MAX}} = O_P(d^{1/2}m_{w,nd}),$$

since $\|\boldsymbol{\Lambda}_H\|_{\mathsf{MAX}} = O(1)$, $\left\|\boldsymbol{\hbar}^\intercal\boldsymbol{w}\right\|_1 \le K_H\left\|\boldsymbol{\hbar}^\intercal\boldsymbol{w}\right\|_{\mathsf{MAX}} = O_P\left((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}}\right)$ by Lemma B.4, $\|\widehat{\boldsymbol{\Lambda}}_H\|_1 \le d^{1/2}\|\widehat{\boldsymbol{\Lambda}}_H\|_F = dK_H$, and $\|\widehat{\boldsymbol{\Lambda}}_H\|_{\mathsf{MAX}} \le \|\widehat{\boldsymbol{\Lambda}}_H\|_F = d^{1/2}K_H$. Therefore

$$\left\|\left(\widehat{\boldsymbol{\Lambda}}_H - \boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R}\right)\widehat{\mathbf{D}}_{x,K_H}\right\|_{\mathsf{MAX}} = O_P\left(d(\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+}} + d^{1/2}m_{w,nd}\right) = O_P(d\cdot a_{nd}). \tag{A.15}$$

Since $\|\widehat{\mathbf{D}}_{x,K_H}^{-1}\|_{\mathsf{sp}} = O_P(d^{-1}n^{-2\underline{\tau}_G^{*-}})$ by Lemma A.1(ii), we can prove the result by noting that

$$n^{-2\underline{\tau}_G^{*-}}\cdot a_{nd} = o_P(1)$$

when $n^{1+4\underline{\tau}_G^{*-}-2\bar{\tau}_V^{*+}-2(\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+})}/(\log d) \to \infty$ and $m_{w,nd}/(d^{1/2}n^{2\underline{\tau}_G^{*-}}) \to 0$.

For part (ii), noting that

$$d^{-1}\widehat{\boldsymbol{\Lambda}}_H^\intercal\widehat{\boldsymbol{\Lambda}}_H = \mathbf{I}_{K_H} \quad\text{and}\quad \left\|\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H\right\|_{\mathsf{sp}} \le (dK_H)^{1/2}\left\|\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H\right\|_{\mathsf{MAX}},$$

we have

$$\begin{aligned}\|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\mathbf{I}_{K_H}\|_{\mathsf{sp}} &= d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H^\intercal\widehat{\mathbf{\Lambda}}_H\|_{\mathsf{sp}}\\ &\leq 2d^{-1}\|\widehat{\mathbf{\Lambda}}_H\|_{\mathsf{sp}}\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\|_{\mathsf{sp}}+d^{-1}\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\|_{\mathsf{sp}}^2\\ &= O_P(n^{-2\underline{\tau}_G^{*-}}\cdot a_{nd})=o_P(1). \end{aligned}\tag{A.16}$$

Then, by triangle inequality, we have $|\|(d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R})\|_{\mathsf{sp}}-1|=o_P(1)$. Since

$$\|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\mathsf{sp}}\|(\mathbf{D}_H\mathbf{R}\mathbf{R}^\intercal\mathbf{D}_H)^{-1}\|_{\mathsf{sp}}\geq\|\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H/d\|_{\mathsf{sp}},\tag{A.17}$$

by Assumption 4, we have $\|(\mathbf{D}_H\mathbf{R}\mathbf{R}^\intercal\mathbf{D}_H)^{-1}\|_{\mathsf{sp}}=\|(\mathbf{D}_H\mathbf{R})^{-1}\|_{\mathsf{sp}}^2=O_P(1)$.

One the other hand, by (A.2) and (A.16),

$$\|(\mathbf{D}_H\mathbf{R})^{-1}(\frac{\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H}{d})^{-1}(\mathbf{R}^\intercal\mathbf{D}_H)^{-1}-\mathbf{I}_{K_H}\|_{\mathsf{sp}}\leq\frac{\|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\mathbf{I}_{K_H}\|_{\mathsf{sp}}}{1-\|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\mathbf{I}_{K_H}\|_{\mathsf{sp}}}=o_P(1).$$

Then following the same argument as in (A.17), we can prove $\|\mathbf{D}_H\mathbf{R}\|_{\mathsf{sp}}=O_P(1)$.

As for (A.12), since $\mathbf{R}^\intercal(\boldsymbol{\hbar}^\intercal\boldsymbol{\hbar})^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}=\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\widehat{\mathbf{\Lambda}}_H$, by (A.10) and (A.16), we have

$$\begin{aligned}&\|d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^\intercal(\boldsymbol{\hbar}^\intercal\boldsymbol{\hbar})^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2}-\mathbf{I}_{K_H}\|_{\mathsf{sp}}\\ &=\|d^{-1}\mathbf{R}^\intercal(\boldsymbol{\hbar}^\intercal\boldsymbol{\hbar})^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}-\mathbf{I}_{K_H}\|_{\mathsf{sp}}\\ &\leq\|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\mathbf{I}_{K_H}\|_{\mathsf{sp}}+d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\|_{\mathsf{sp}}\|\widehat{\mathbf{\Lambda}}_H-\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\mathsf{sp}}\\ &=O_P(n^{-2\underline{\tau}_G^{*-}}\cdot a_{nd})=o_P(1).\end{aligned}$$

Then by Lemmas B.1 and B.2, we can prove $\|\boldsymbol{\hbar}^\intercal\boldsymbol{\hbar}\|_{\mathsf{sp}}=O_P(1)$, and therefore we have (A.12).

For part (iii), we use the following decomposition

$$\widehat{\boldsymbol{\hbar}}^\intercal-\mathbf{R}^{-1}\boldsymbol{\hbar}^\intercal=d^{-1}\widehat{\mathbf{\Lambda}}_H^\intercal\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\right)\mathbf{R}^{-1}\boldsymbol{\hbar}^\intercal-d^{-1}\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\right)^\intercal\boldsymbol{w}^\intercal+d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\boldsymbol{w}^\intercal.\tag{A.18}$$

For the first term on the right hand side (RHS) of (A.18), we have

$$\begin{aligned}&\|\widehat{\mathbf{\Lambda}}_H^\intercal\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\right)\mathbf{R}^{-1}\boldsymbol{\hbar}^\intercal\|_{\mathsf{sp}}\\ &\leq\|\widehat{\mathbf{\Lambda}}_H^\intercal\|\left\|\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\right)\widehat{\mathbf{D}}_{x,K_H}\right\|_{\mathsf{sp}}\left\|\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\right\|_{\mathsf{sp}}\left\|(\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2})^{-1}\right\|_{\mathsf{sp}}\|\boldsymbol{\hbar}^\intercal\|_{\mathsf{sp}}\\ &=O_P(d^{1/2})\cdot O_P(d\cdot a_{nd})\cdot O_P(d^{-1/2}n^{-\underline{\tau}_G^{*-}})\cdot O_P(1)\cdot O_P(1)\\ &=O_P(dn^{-\underline{\tau}_G^{*-}}a_{nd}).\end{aligned}\tag{A.19}$$

For the second term on the RHS of (A.18), when $n^{1+4\underline{\tau}_G^{*-}-4\bar{\tau}_V^{*+}}/\log d \to \infty$, we have

$$
\begin{aligned}
&\left\|\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\right)^\intercal \boldsymbol{w}\right\|_{\mathsf{sp}} \le \left\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}-\widehat{\mathbf{\Lambda}}_H\right\|_{\mathsf{sp}}\|\boldsymbol{w}\|_{\mathsf{sp}}\\
&= O_P(d^{1/2}n^{-2\underline{\tau}_G^{*-}}a_{nd})\cdot O_P(d^{1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^{*+}}+m_{w,nd}^{1/2})\\
&= O_P(dn^{-\underline{\tau}_G^{*-}}a_{nd})\cdot O_P(n^{-\underline{\tau}_G^{*-}}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^{*+}}+m_{w,nd}^{1/2}d^{-1/2})\\
&= o_P(dn^{-\underline{\tau}_G^{*-}}a_{nd}).
\end{aligned}
\tag{A.20}
$$

For the last term on the RHS of (A.18), by Lemma B.4(v), we have

$$
\begin{aligned}
d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\boldsymbol{w}^\intercal\|_{\mathsf{sp}} &\le d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\|_{\mathsf{sp}}\|\mathbf{\Lambda}_H^\intercal\boldsymbol{w}^\intercal\|_{\mathsf{sp}}\\
&= d^{-1}\cdot O_P(1)\cdot O_P\left(d^{1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^{*+}}m_{w,nd}^{1/2}+d^{1/2}m_{w,nd}^{1/2}\right)\\
&= o_P(n^{-\underline{\tau}_G^{*-}}a_{nd}),
\end{aligned}
\tag{A.21}
$$

when $d^{-1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^{*+}}m_{w,nd}^{1/2}=o((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+}-\underline{\tau}_G^{*-}})$, or equivalently, $n^{1-4(\bar{\tau}_G^{*+}\vee\bar{\tau}_V^{*+})+4\underline{\tau}_G^{*-}}=o(d^2\log d/m_{w,nd}^2)$. Combing (A.19)–(A.21), we have $\|\widehat{\hbar}^\intercal-\mathbf{R}^{-1}\hbar^\intercal\|_{\mathsf{sp}}=O_P(n^{-\underline{\tau}_G^{*-}}a_{nd})$, which completes the proof of Lemma A.2. □

Proof of Theorem 3.1. (i) Note that $\widehat{\hbar}^*=\widehat{\hbar}(\widehat{\hbar}^\intercal\widehat{\hbar})^{-1/2}=d^{1/2}\widehat{\hbar}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}$ and

$$
(\mathbf{R}^*)^{-1}=d^{1/2}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\mathbf{R}^{-1}(\hbar^\intercal\hbar)^{1/2}.
\tag{A.22}
$$

By Lemma A.2(iii), we have

$$
\begin{aligned}
\left\|\widehat{\hbar}^{*\intercal}-(\mathbf{R}^*)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_{\mathsf{MAX}} &= \left\|\widehat{\hbar}^{*\intercal}-d^{1/2}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\mathbf{R}^{-1}(\hbar^\intercal\hbar)^{1/2}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_{\mathsf{MAX}}\\
&= \left\|\widehat{\hbar}^{*\intercal}-d^{1/2}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\mathbf{R}^{-1}\hbar^\intercal\right\|_{\mathsf{MAX}}\\
&\le d^{1/2}\|\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\|_{\mathsf{sp}}\|\widehat{\hbar}^\intercal-\mathbf{R}^{-1}\hbar^\intercal\|_{\mathsf{sp}}\\
&= O_P\left(n^{-2\underline{\tau}_G^{*-}}\cdot a_{nd}\right).
\end{aligned}
$$

(ii) Following (A.15) and noting that $\widehat{\mathbf{\Lambda}}_H^*=\widehat{\mathbf{\Lambda}}_H(\widehat{\hbar}^\intercal\widehat{\hbar})^{1/2}=d^{-1/2}\widehat{\mathbf{\Lambda}}_H\widehat{\mathbf{D}}_{x,K_H}^{1/2}$, we have

$$
\left\|\widehat{\mathbf{\Lambda}}_H^*-d^{-1/2}\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\right\|_{\mathsf{MAX}}=O_P\left(n^{-2\underline{\tau}_G^{*-}}\cdot a_{nd}\right).
$$

Using the notation of $\mathbf{R}^*$, it can be equivalently written as

$$\left\|\widehat{\mathbf{\Lambda}}_H^* - \mathbf{\Lambda}_H\mathbf{D}_H(\hbar^\intercal\hbar)^{1/2}\mathbf{R}^*\right\|_{\mathsf{MAX}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right).$$

(iii) For the first part, it is obvious that $\widehat{\hbar}^*\widehat{\mathbf{\Lambda}}_H^{*\intercal} = \widehat{\hbar}\widehat{\mathbf{\Lambda}}_H^\intercal$. For the second part, we have

$$\begin{aligned}
&\left\|\widehat{\hbar}^*\widehat{\mathbf{\Lambda}}_H^{*\intercal} - \hbar\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\right\|_{\mathsf{MAX}} \\
&\le \left\|\widehat{\hbar}^{*\intercal} - (\mathbf{R}^*)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_1 \|\mathbf{\Lambda}_H\|_{\mathsf{MAX}} \left\|\mathbf{D}_H(\hbar^\intercal\hbar)^{1/2}\mathbf{R}^*\right\|_1 \\
&+ \left\|(\mathbf{R}^*)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_1 \left\|\widehat{\mathbf{\Lambda}}_H^* - \mathbf{\Lambda}_H\mathbf{D}_H(\hbar^\intercal\hbar)^{1/2}\mathbf{R}^*\right\|_{\mathsf{MAX}} \\
&+ \left\|\widehat{\hbar}^{*\intercal} - (\mathbf{R}^*)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_1 \left\|\widehat{\mathbf{\Lambda}}_H^* - \mathbf{\Lambda}_H\mathbf{D}_H(\hbar^\intercal\hbar)^{1/2}\mathbf{R}^*\right\|_{\mathsf{MAX}} \\
&= O_P\left(n^{\bar{\tau}_G^{*+} - 2\underline{\tau}_G^{*-}} \cdot a_{nd}\right),
\end{aligned}$$

as $\|\mathbf{\Lambda}_H\|_{\mathsf{MAX}} = O(1)$, $\left\|\mathbf{D}_H(\hbar^\intercal\hbar)^{1/2}\mathbf{R}^*\right\|_1 = O_P(n^{\bar{\tau}_G^{*+}})$, and

$$\left\|\widehat{\hbar}^{*\intercal} - (\mathbf{R}^*)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_1 \le K_H^{1/2}\left\|\widehat{\hbar}^{*\intercal} - (\mathbf{R}^*)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_{\mathsf{sp}} = O_P(n^{\bar{\tau}_G^{*+}} \cdot a_{nd}).$$

Thus, we obtain the uniformly convergence rate for the common components.

(iv) We next prove that $\mathbf{R}^*$ is an asymptotically orthogonal matrix. By (A.22), we have

$$\mathbf{R}^{*\intercal}\mathbf{R}^* = d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^\intercal(\hbar^\intercal\hbar)^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2} = d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\widehat{\mathbf{\Lambda}}_H\widehat{\mathbf{D}}_{x,K_H}^{-1/2}.$$

Thus by (A.10) and (A.16), we have

$$\begin{aligned}
&\|\mathbf{R}^{*\intercal}\mathbf{R}^* - \mathbf{I}_{K_H}\|_{\mathsf{sp}} \\
&\le \|d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\widehat{\mathbf{\Lambda}}_H\widehat{\mathbf{D}}_{x,K_H}^{-1/2} - d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\|_{\mathsf{sp}} \\
&+ \|d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{-1/2} - \mathbf{I}_{K_H}\|_{\mathsf{sp}} \\
&= \|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\widehat{\mathbf{\Lambda}}_H - d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\mathsf{sp}} \\
&+ \|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \mathbf{I}_{K_H}\|_{\mathsf{sp}} \\
&= O_P(n^{-2\underline{\tau}_G^{*-}}a_{nd}).
\end{aligned} \tag{A.23}$$

We thus complete the proof of Theorem 3.1. □

To prove Lemma 3.1, we need some intermediate estimators or infeasible estimators related to $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_\perp$. Recall that $\widehat{\boldsymbol{\beta}}_\perp$ is the matrix of eigenvectors associated with the largest K_F eigenvalues of

$\widehat{\mathbf{S}}_{HH} := n^{-1}\widehat{\mathscr{H}}^{*c\intercal}\widehat{\mathscr{H}}^{*c}$, and that $\widehat{\boldsymbol{\beta}}$ is the matrix of eigenvectors associated with the rest of the K_G eigenvalues. For a $K_H \times K_H$ matrix, Ξ, we define $\mathbf{S}^{\Xi}_{HH} := n^{-1}\Xi\mathscr{H}^{c\intercal}\mathscr{H}^{c}\Xi^\intercal$, where $\mathscr{H}^c = \mathscr{H} - \overline{\mathscr{H}}$ and $\overline{\mathscr{H}} = n^{-1}\mathbf{1}_n\sum_{s=1}^n \boldsymbol{H}^\intercal_{s\Delta}$. Replacing $\widehat{\mathbf{S}}_{HH}$ with $\mathbf{S}^{\Xi}_{HH}$, we can obtain the infeasible estimators, $\boldsymbol{\beta}^{\Xi}_\perp$ and $\boldsymbol{\beta}^{\Xi}$. For simplicity, we use $\mathbf{S}_{HH}$ to denote $\mathbf{S}^{\mathbf{I}_{K_H}}_{HH}$. Later on, we will determine a proper choice of Ξ.

Lemma A.3. *Suppose that Assumptions 1, 3 and 6 are satisfied. If the eigenvalues of $\Xi\Xi^\intercal$ are bounded away from zero and infinity uniformly with probability approaching one, then $\boldsymbol{\beta}^{\Xi}_\perp$ and $\boldsymbol{\beta}^{\Xi}$ are super-consistent in the sense that*

$$\boldsymbol{\beta}^{\Xi} - \Xi^\intercal\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\Xi\Xi^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\Xi\boldsymbol{\beta}^{\Xi} = O_P(n^{-1}), \tag{A.24}$$

and

$$\boldsymbol{\beta}^{\Xi}_\perp - \Xi^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\boldsymbol{\beta}^{\Xi}_\perp = O_P(n^{-1}). \tag{A.25}$$

Proof. We decompose $\boldsymbol{\beta}^{\Xi}$ in the directions of $\Xi^\intercal\boldsymbol{\beta}$ and $\Xi^{-1}\boldsymbol{\beta}_\perp$ (which are orthogonal) as

$$\begin{aligned}\boldsymbol{\beta}^{\Xi} \;=\; & \Xi^\intercal\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\Xi\Xi^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\Xi\boldsymbol{\beta}^{\Xi} \\ & +\Xi^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\boldsymbol{\beta}^{\Xi}.\end{aligned} \tag{A.26}$$

Note that $\boldsymbol{\beta}^{\Xi}$ satisfies $\Xi^{-1}\mathbf{S}_{HH}(\Xi^\intercal)^{-1}\boldsymbol{\beta}^{\Xi} = \boldsymbol{\beta}^{\Xi}\mathbf{D}^{\Xi}_S$, where $\mathbf{D}^{\Xi}_S$ is a $K_G \times K_G$ diagonal matrix with the diagonal elements being the K_G smallest eigenvalues of $\Xi^{-1}\mathbf{S}_{HH}(\Xi^\intercal)^{-1}$ arranged in a descending order. Using (A.26) and the equality $\mathbf{I}_{K_H} = \boldsymbol{\beta}_\perp\boldsymbol{\beta}^\intercal_\perp + \boldsymbol{\beta}\boldsymbol{\beta}^\intercal$, we have

$$\begin{aligned}\boldsymbol{\beta}^\intercal_\perp\Xi^{-1}\boldsymbol{\beta}^{\Xi}\mathbf{D}^{\Xi}_S =& \boldsymbol{\beta}^\intercal_\perp\Xi\Xi^{-1}\mathbf{S}_{HH}(\Xi^\intercal)^{-1}\boldsymbol{\beta}^{\Xi} \\ =& \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\Xi\Xi^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\Xi\boldsymbol{\beta}^{\Xi} \\ & + \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\mathbf{I}_{K_H}(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}^\intercal_\perp\Xi^{-1}\boldsymbol{\beta}^{\Xi} \\ =& \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\Xi\Xi^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\Xi\boldsymbol{\beta}^{\Xi} \\ & + \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}^\intercal_\perp\Xi^{-1}\boldsymbol{\beta}^{\Xi} \\ & + \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}_\perp\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}^\intercal_\perp\Xi^{-1}\boldsymbol{\beta}^{\Xi}, \\ =& \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\Xi\Xi^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\Xi\boldsymbol{\beta}^{\Xi} \\ & + \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}^\intercal_\perp\Xi^{-1}\boldsymbol{\beta}^{\Xi} \\ & + \boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}_\perp\boldsymbol{\beta}^\intercal_\perp\Xi^{-1}\boldsymbol{\beta}^{\Xi}.\end{aligned}$$

Vectorizing this expression, we have

$$\begin{aligned}\mathsf{vec}(\boldsymbol{\beta}^\intercal_\perp\Xi^{-1}\boldsymbol{\beta}^{\Xi}) \;=\; & \big\{\mathbf{D}^{\Xi}_S\otimes\mathbf{I}_{K_F} - \mathbf{I}_{K_G}\otimes\boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}_\perp \\ & -\mathbf{I}_{K_G}\otimes\boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}^\intercal_\perp(\Xi^\intercal)^{-1}\Xi^{-1}\boldsymbol{\beta}_\perp]^{-1}\big\}^{-1} \\ & \cdot\mathsf{vec}\left(\boldsymbol{\beta}^\intercal_\perp\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\Xi\Xi^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\Xi\boldsymbol{\beta}^{\Xi}\right).\end{aligned} \tag{A.27}$$

Recall that $\boldsymbol{\beta} = (\mathbf{O}_{K_G\times K_F} \quad \mathbf{I}_{K_G})^\intercal$ and $\boldsymbol{\beta}_\perp = (\mathbf{I}_{K_F} \quad \mathbf{O}_{K_F\times K_G})^\intercal$. By Lemma B.5, we have

$$
\begin{aligned}
\boldsymbol{\beta}^\intercal \mathbf{S}_{HH}\boldsymbol{\beta} &= n^{-2}\sum_{s=1}^{n} \boldsymbol{G}^c_{s\Delta}\boldsymbol{G}^{c\intercal}_{s\Delta} = O_P(n^{-1}),\\
\boldsymbol{\beta}_\perp^\intercal \mathbf{S}_{HH}\boldsymbol{\beta}_\perp &= n^{-1}\sum_{s=1}^{n} \boldsymbol{F}^c_{s\Delta}\boldsymbol{F}^{c\intercal}_{s\Delta} \text{ is bounded away from zero,}\\
\boldsymbol{\beta}^\intercal \mathbf{S}_{HH}\boldsymbol{\beta}_\perp &= n^{-3/2}\sum_{s=1}^{n} \boldsymbol{G}^c_{s\Delta}\boldsymbol{F}^{c\intercal}_{s\Delta} = O_P(n^{-1}),
\end{aligned}
$$

where $\boldsymbol{G}^c_{s\Delta} = \boldsymbol{G}_{s\Delta} - n^{-1}\sum_{s=1}^{n}\boldsymbol{G}_{s\Delta}$ and $\boldsymbol{F}^c_{s\Delta} = \boldsymbol{F}_{s\Delta} - n^{-1}\sum_{s=1}^{n}\boldsymbol{F}_{s\Delta}$. Thus, only the first block, $\boldsymbol{\beta}_\perp^\intercal \mathbf{S}_{HH}\boldsymbol{\beta}_\perp$, of the matrix $\mathbf{S}_{HH} = \begin{bmatrix} \boldsymbol{\beta}_\perp^\intercal \mathbf{S}_{HH}\boldsymbol{\beta}_\perp & \boldsymbol{\beta}_\perp^\intercal \mathbf{S}_{HH}\boldsymbol{\beta} \\ \boldsymbol{\beta}^\intercal \mathbf{S}_{HH}\boldsymbol{\beta}_\perp & \boldsymbol{\beta}^\intercal \mathbf{S}_{HH}\boldsymbol{\beta} \end{bmatrix}$ does not converge to zero. Therefore, $\mathbf{D}^{\boldsymbol{\Xi}}_S = o_P(1)$, and we have $\boldsymbol{\beta}_\perp^\intercal \boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}} = O_P(n^{-1})$. Then using (A.26) again, we can prove the consistency of $\boldsymbol{\beta}^{\boldsymbol{\Xi}}$. Using the same argument, we can prove the consistency of $\boldsymbol{\beta}^{\boldsymbol{\Xi}}_\perp$. $\square$

When $\boldsymbol{\Xi} = \mathbf{I}_{K_H}$, the results degenerate to Lemma 1 of Harris (1997). When $\boldsymbol{\Xi} = (\hbar^\intercal \hbar)^{1/2}(\mathbf{R}^*)$, and replacing $\mathbf{S}_{HH}$ with $\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal$, we can prove Lemma 3.1.

Proof of Lemma 3.1. Note that $\widehat{\boldsymbol{\beta}}$ satisfies $\widehat{\mathbf{S}}_{HH}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}\widehat{\mathbf{D}}_S$, where $\widehat{\mathbf{D}}_S$ is a $K_G \times K_G$ diagonal matrix with the diagonal elements being the K_G smallest eigenvalues of $\widehat{\mathbf{S}}_{HH}$ arranged in a descending order. Let $\boldsymbol{\Xi} = (\hbar^\intercal \hbar)^{1/2}(\mathbf{R}^*)^\intercal$. Following similar arguments in the proof of Lemma A.3, we have

$$
\begin{aligned}
\mathsf{vec}(\boldsymbol{\beta}_\perp^\intercal \boldsymbol{\Xi}^{-1}\widehat{\boldsymbol{\beta}}) = &\Big\{\widehat{\mathbf{D}}^{\boldsymbol{\Xi}}_S \otimes \mathbf{I}_{K_F} - \mathbf{I}_{K_G}\otimes \boldsymbol{\beta}_\perp^\intercal[\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal]\boldsymbol{\beta}_\perp \\
&-\mathbf{I}_{K_G}\otimes \boldsymbol{\beta}_\perp^\intercal[\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal]\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\Big\}^{-1}\\
&\cdot \mathsf{vec}\left(\boldsymbol{\beta}_\perp^\intercal[\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal]\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\widehat{\boldsymbol{\beta}}\right). \qquad \text{(A.28)}
\end{aligned}
$$

Using the convergence results in Lemma B.6 and $\|\widehat{\mathbf{D}}_S\|_{\mathsf{sp}} \le \|\mathbf{D}^{\boldsymbol{\Xi}}_S\|_{\mathsf{sp}} + \|\widehat{\mathbf{S}}_{HH} - \boldsymbol{\Xi}^{-1}\mathbf{S}_{HH}(\boldsymbol{\Xi}^\intercal)^{-1}\|_{\mathsf{sp}} = o_P(1)$, we can prove $\boldsymbol{\beta}_\perp^\intercal \boldsymbol{\Xi}^{-1}\widehat{\boldsymbol{\beta}} = O_P(n^{-1})$. Then following the same arguments as in the proof of Lemma A.3, we can prove the results. $\square$

Proof of Theorem 3.2. (ii) and (iii) follow directly from Theorem 3.1 and Lemma 3.1 by noting that $a_{nd} > n^{-1}$.

As for (i), by Theorem 3.1 and Lemma 3.1, we have

$$
\left\|\widehat{\hbar}^*\widehat{\boldsymbol{\beta}}_\perp - \hbar(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp\mathbf{Q}_{\boldsymbol{\beta}_\perp}\right\|_{\mathsf{sp}} = O_P\left(n^{-2\tau^{*-}_G}\cdot a_{nd}\right). \qquad \text{(A.29)}
$$

Using $(\boldsymbol{\beta}_\perp\boldsymbol{\beta}_\perp^\intercal + \boldsymbol{\beta}\boldsymbol{\beta}^\intercal) = \mathbf{I}_{K_H}$, we have

$$\begin{aligned}\hbar(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp\mathbf{Q}_{\boldsymbol{\beta}_\perp} &= \hbar(\boldsymbol{\beta}_\perp\boldsymbol{\beta}_\perp^\intercal + \boldsymbol{\beta}\boldsymbol{\beta}^\intercal)(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp \\ &= \ell\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp + g\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp.\end{aligned}$$

Therefore, we only need to prove that the second term on the RHS of the second equality above is $O_P(n^{-2\underline{\tau}_G^{*-}}a_{nd})$. Indeed,

$$\begin{aligned}\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp &= \boldsymbol{\beta}^\intercal(\hbar^\intercal\hbar)^{-1/2}(\mathbf{R}^{*\intercal}\mathbf{R}^*)^{-1}(\hbar^\intercal\hbar)^{-1/2}\boldsymbol{\beta}_\perp \\ &= \boldsymbol{\beta}^\intercal(\hbar^\intercal\hbar)^{-1}\boldsymbol{\beta}_\perp + O_P(n^{-2\underline{\tau}_G^{*-}}a_{nd}) = O_P(n^{-2\underline{\tau}_G^{*-}}a_{nd}),\end{aligned}$$

where the last two equalities follow from (A.23) and the result that $\hbar^\intercal\hbar$ converges to a block diagonal matrix at rate $(\log d/n)^{1/2}$ using Lemmas B.1–B.3. Thus we complete the proof of (i).

As for (iv), by Theorem 3.1 and Lemma 3.1, we have

$$\left\|\widehat{\boldsymbol{\Lambda}}_H^*\widehat{\boldsymbol{\beta}} - \boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}\mathbf{Q}_{\boldsymbol{\beta}}\right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}}\cdot a_{nd}\right). \tag{A.30}$$

Following similar arguments to the proof of part (i), we can show that $\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta} = O_P(n^{-2\underline{\tau}_G^{*-}}a_{nd})$ and that

$$\left\|\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}\mathbf{Q}_{\boldsymbol{\beta}} - \boldsymbol{\Lambda}_G\mathbf{D}_G\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\widehat{\boldsymbol{\beta}}\right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}}\cdot a_{nd}\right). \tag{A.31}$$

Combining (A.30) and (A.31), we complete the proof. □

Proof of Theorem 3.3.
Since $\|\mathbf{L}_n\|_{\mathsf{sp}} = 1$, using the submultiplicative property of the spectral norm, the convergence rate of the estimators for the first-differenced factors is also the convergence rate of the estimators for the cumulated factors. □

Appendix B: Auxiliary Lemmas

Recall that $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ are increments of continuous-time processes, between t and $t-\Delta$, while $\boldsymbol{g}_t$ and $\boldsymbol{v}_t$ are the first-order differences of stationary time series, $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$, for $t = 0, \Delta, \ldots, n\Delta$. Lemmas B.1–B.3 give the large deviation theory for them. Specially, Lemma B.1 is for $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ only, Lemma B.2 for $\boldsymbol{g}_t$ and $\boldsymbol{v}_t$ only, and Lemma B.3 for both the continuous-time processes and the discrete-time processes.

Lemma B.1. *Under Assumption 1, we have*
(i) $\left\|\sum_{s=1}^n \boldsymbol{u}_{s\Delta}\boldsymbol{u}_{s\Delta}^\intercal - \boldsymbol{\Sigma}_U\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$;
(ii) $\left\|\sum_{s=1}^n \boldsymbol{f}_{s\Delta}\boldsymbol{f}_{s\Delta}^\intercal - \boldsymbol{\Sigma}_F\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$;
(iii) $\left\|\sum_{s=1}^n \boldsymbol{u}_{s\Delta}\boldsymbol{f}_{s\Delta}^\intercal\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$.
(iv) In addition, if Assumptions 4 and 5 hold, we have
$\left\|d^{-1}\sum_{s=1}^n \boldsymbol{\Lambda}_H^\intercal\boldsymbol{u}_{s\Delta}\boldsymbol{u}_{s\Delta}^\intercal\boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\right\|_{\mathsf{MAX}} = O_P(m_{U,d}(\log d/n)^{1/2})$.

Proof. Parts (i)–(iii) are the same as Lemma 1 in Aït-Sahalia and Xiu (2017). We only prove part (iv), as parts (i)–(iii) can be proved similarly. By Bonferroni inequality and Lemma 10 of Tao et al. (2013b), we have

$$P\left(\left\|\sum_{s=1}^n d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{u}_{s\Delta}\boldsymbol{u}_{s\Delta}^\intercal\boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\right\|_{\mathsf{MAX}} > c\right) \le K_h^2 \cdot 4\exp(-nc^2/(64C_1))$$

for all $0 \le c \le \mu_d^2(d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H)\cdot n^{1/2}$, where $C_1 = 8\|d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\|_{\mathsf{MAX}}^2$ is obtained from Lemma 3 of Fan et al. (2012). By Assumptions 1 and 4, we have that $\mu_d^2(d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H)$ is bounded away from zero, and $C_1 \le 8\|d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\|_{\mathsf{sp}}^2 = O(m_{U,d}^2)$. Then using the exponential inequality and taking $x = m_{U,d}(\log d/n)^{1/2}$, we can prove the result. □

Lemma B.2. *Under Assumption 6, we have*
(i) $\left\|n^{-1}\sum_{s=1}^n \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal - \boldsymbol{\Sigma}_v\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$;
(ii) $\left\|n^{-1}\sum_{s=1}^n \boldsymbol{g}_{s\Delta}\boldsymbol{g}_{s\Delta}^\intercal - \boldsymbol{\Sigma}_g\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$;
(iii) $\left\|n^{-1}\sum_{s=1}^n \boldsymbol{v}_{s\Delta}\boldsymbol{g}_{s\Delta}^\intercal\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$.
(iv) In addition, if Assumptions 4 and 5 hold, we have
$\left\|(nd)^{-1}\sum_{s=1}^n \boldsymbol{\Lambda}_H^\intercal\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal\boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_v\boldsymbol{\Lambda}_H\right\|_{\mathsf{MAX}} = O_P(m_{v,d}(\log d/n)^{1/2})$.

Proof. Parts (i)–(iii) are the same as Lemma C.3 in Fan et al. (2013). Note that, under Assumption 4(i), the mixing coefficient of $\{\boldsymbol{v}_{s\Delta}\}$ is bounded by $C'_\alpha\exp(-s^{\gamma_2})$, for some positive constant C'_α. Also note that $v_{i,s\Delta}$ still satisfies the exponential-type tail condition, since

$$\begin{aligned}\max_{1\le i\le d} P\left(|v_{i,s\Delta}| > c\right) &\le \max_{1\le i\le d} 2P\left(|V_{i,s\Delta}| > c/2\right) \\ &\le 2\exp(1-(c/(2b_1))^{\gamma_2}) \le \exp(1-(c/b_3)^{\gamma_4}), \qquad \text{(B.1)}\end{aligned}$$

for $1 \le i \le d$, $s = 1, \cdots, n$, and $c > 0$, where $\gamma_4 \in (0, \gamma_2)$ and $b_3 > 2b_1 \max\{(\gamma_4/\gamma_2)^{1/\gamma_2}, (1+\log 2)^{1/\gamma_2}\}$, and the last inequality is shown in the proof of Lemma C.2 of Fan et al. (2011). Again by Lemma C.2 of Fan et al. (2011), $|v_{i_1,s\Delta} v_{i_2,s\Delta}|$ still satisfies the exponential-type tail condition,

$$\max_{1 \le i \le d} P\left(|v_{i_1,s\Delta} v_{i_2,s\Delta} - \mathsf{E}[v_{i_1,s\Delta} v_{i_2,s\Delta}]| > c\right) \le \exp(1 - (c/b_4)^{\gamma_5}), \tag{B.2}$$

for $1 \le i_1, i_2 \le d$, $s = 1, \cdots, n$, $c > 0$, some b_4, and $\gamma_5 \in (0, \gamma_4/2)$. Therefore, using the arguments in the proof of Lemma A.3 in Fan et al. (2011), we can show that

$$P\left(\left\|n^{-1}\sum_{s=1}^{n} \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_v\right\|_{\mathsf{MAX}} > C_2\sqrt{\frac{\log d}{n}}\right) = O\left(\frac{1}{d^2}\right)$$

for some positive constant C_2, which proves part (i). Parts (ii) and (iii) are similar to part (i) and can be obtained from the inequalities derived in Lemma B.1 of Fan et al. (2011). As for part (iv), we have

$$\begin{aligned} & P\left(\left\|n^{-1}\sum_{s=1}^{n} \boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\Lambda}_H - \boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_v\boldsymbol{\Lambda}_H\right\|_{\mathsf{MAX}} > dm_{v,d}\cdot x\right) \\ \le\ & K_H^2 \max_{1\le j_1,j_2\le K_H} P\left(\left|n^{-1}\sum_{s=1}^{n} \boldsymbol{\lambda}_{H,j_1}^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\lambda}_{H,j_2} - \boldsymbol{\lambda}_{H,j_1}^{\intercal}\boldsymbol{\Sigma}_v\boldsymbol{\lambda}_{H,j_2}\right| > dm_{v,d}\cdot x\right). \end{aligned} \tag{B.3}$$

Applying similar arguments in (B.1) and (B.2) and using Lemma C.2 of Fan et al. (2011) under Assumption 5(iv), we have

$$\max_{1\le j_1,j_2\le K_H} P\left((dm_{v,d})^{-1}|\boldsymbol{\lambda}_{H,j_1}^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\lambda}_{H,j_2} - \mathsf{E}[\boldsymbol{\lambda}_{H,j_1}^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\lambda}_{H,j_2}]| > c\right) \le \exp(1 - (c/b_5)^{\gamma_6}), \tag{B.4}$$

for $\gamma_6 \in (0, \gamma_2\gamma_3/(\gamma_2+\gamma_3))$, $c > 0$, and some $b_5 > 0$ which does not depend on n and d. Since $|\boldsymbol{\lambda}_{H,j_1}^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\lambda}_{H,j_2}|$ satisfies the strong mixing condition, we can follow the same arguments as the proof of Lemma B.1 in Fan et al. (2011) by applying the Bernstein's inequality in Theorem 1 of Merlevède et al. (2011) to obtain

$$P\left((dm_{v,d})^{-1}\left|n^{-1}\sum_{s=1}^{n} \boldsymbol{\lambda}_{H,j_1}^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\lambda}_{H,j_2} - \boldsymbol{\lambda}_{H,j_1}^{\intercal}\boldsymbol{\Sigma}_v\boldsymbol{\lambda}_{H,j_2}\right| > C_3\sqrt{\frac{\log d}{n}}\right) = O\left(\frac{1}{d^2}\right), \tag{B.5}$$

for some positive constant C_3, which only depends on γ_1, γ_6 and b_5. Then by (B.3) and (B.5), we can complete the proof of part (iv). □

Lemma B.3. *Under Assumptions 1, 3 and 6, we have*
(i) $\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$*;*
(ii) $\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{g}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$*;*
(iii) $\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{v}_{s\Delta}\boldsymbol{f}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2})$*;*
(iv) $\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{g}_{s\Delta}\boldsymbol{f}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} = O_P((1/n)^{1/2})$*;*
(v) $\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{G}_{s\Delta}(n^{-1/2}\boldsymbol{F}_{s\Delta}^{\intercal})\right\|_{\mathsf{MAX}} = O_P((1/n)^{1/2})$*.*

Proof. (i) The proof is similar to that of Lemma 11 in Tao et al. (2013b). Since $\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal} = \sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal} - \sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{(s-1)\Delta}^{\intercal}$, we only need to prove

$$n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal} = O_P((\log d/n)^{1/2}) \tag{B.6}$$

and

$$n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{(s-1)\Delta}^{\intercal} = O_P((\log d/n)^{1/2}). \tag{B.7}$$

The proofs of (B.6) and (B.7) are similar, so we only provide the former. Denote

$$\Omega_0 = \{\max_{1\le i\le d}\max_{1\le s\le n}|u_{i,s\Delta}| \le 1\}.$$

Using the Bonferroni and Markov inequalities, we have

$$\mathsf{P}(\Omega_0^c) \;=\; \mathsf{P}\left(\max_{1\le i\le d}\max_{1\le s\le n}|u_{i,s\Delta}| > 1\right) \le nde^{C_\sigma/2-n}. \tag{B.8}$$

Note that $\boldsymbol{U}_t$ and $\boldsymbol{V}_t$ are independent. Conditional on the whole path of $\boldsymbol{U}_t$, we have

$$\begin{aligned}
&\mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} > c(\log d/n)^{1/2}\right)\\
\le\;& \mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} > c(\log d/n)^{1/2}, \Omega_0\right) + \mathsf{P}\left(\Omega_0^c\right)\\
\le\;& \mathsf{E}\left[\mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} > c(\log d/n)^{1/2}, \Omega_0\middle|\boldsymbol{U}_t,\, t\in[0,1]\right)\right] + O(nde^{-n})\\
=\;& \mathsf{E}\left[\mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} > c(\log d/n)^{1/2}\middle|\Omega_0, \boldsymbol{U}_t,\, t\in[0,1]\right)\right] + O(nde^{-n}).
\end{aligned} \tag{B.9}$$

Note that conditional on the path of $\boldsymbol{U}_t$ and Ω_0, $\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}$ satisfies the same mixing condition and exponential-tail condition for $\boldsymbol{V}_{s\Delta}$, and the coefficients in these conditions only depend on γ_1, γ_2, and b_1. Thus we can apply the Bernstein's inequality in Theorem 1 of Merlevède et al. (2011) to obtain (letting $\bar{c} = c(\log d/n)^{1/2}$)

$$
\begin{aligned}
& \mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} > \bar{c}\,\middle|\,\Omega_0, \boldsymbol{U}_t,\, t\in[0,1]\right) \\
\leq\ & nd^2\exp\left(-\frac{\bar{c}^{\gamma}}{C_4}\right) + d^2\exp\left(-\frac{\bar{c}^2}{C_5(1+C_6 n)}\right) + d^2\exp\left(-\frac{\bar{c}^2}{C_7 n}\exp\left(\frac{\bar{c}^{\gamma(1-\gamma)}}{C_8((\log\bar{c})^{\gamma})}\right)\right) \\
=\ & O(1/d^2), \qquad \text{(B.10)}
\end{aligned}
$$

when $(\log d)^{2/\gamma-1} = o(n)$ and c is large enough, where $\gamma = 1/\gamma_1 + 1/\gamma_2$, and C_4–C_8 only depends on γ_1, γ_2, and b_1. Therefore (B.10) holds true uniformly for all path of $\boldsymbol{U}_t$ satisfying Ω_0. Combining (B.9) and (B.10), we can prove (B.6). The proofs of (ii)–(v) are similar to that of (i) by choosing proper $\bar{c}$. So we omit them to save space. □

Lemma B.4. *Under Assumptions 1–6, we have*
(i) $\left\|\boldsymbol{w}^{\intercal}\boldsymbol{w} - \boldsymbol{\Sigma}_w\right\|_{\mathsf{MAX}} = O_P((\log d/n)^{1/2}\cdot n^{2\bar{\tau}_V^{*+}})$*;*
(ii) $\left\|\boldsymbol{w}^{\intercal}\boldsymbol{\hbar}\mathbf{D}_H\right\|_{\mathsf{MAX}} = O_P\left((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}}\right)$*;*
(iii) $\left\|\mathbf{D}_H\boldsymbol{\hbar}^{\intercal}\boldsymbol{w}\right\|_1 = O_P\left((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{*+}+\bar{\tau}_G^{*+}}\right)$*;*
(iv) $\|\boldsymbol{w}\|_{\mathsf{sp}} = O_P(d^{1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^{*+}} + m_{U,d}^{1/2} + n^{\bar{\tau}_V^{*}}m_{v,d}^{1/2})$*;*
(v) $\|\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{w}^{\intercal}\|_{\mathsf{sp}} = O_P(d^{1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^{*+}} + d^{1/2}(m_{U,d}^{1/2} + n^{\bar{\tau}_V^{*}}m_{v,d}^{1/2}))$*;*
where $\bar{\tau}_G^{*+} = (1/2+\bar{\tau}_G)_+$*,* $\bar{\tau}_V^{*} = (1/2+\bar{\tau}_V)$*, and* $\bar{\tau}_V^{*+} = (1/2+\bar{\tau}_V)_+$*.*

Proof. For part (i), recall that $\boldsymbol{w}\boldsymbol{w}^{\intercal} = \sum_{s=1}^{n}(\boldsymbol{u}_{s\Delta} + \mathbf{D}_V\boldsymbol{v}_{s\Delta})^{\intercal}(\boldsymbol{u}_{s\Delta} + \mathbf{D}_V\boldsymbol{v}_{s\Delta})$ and $\boldsymbol{\Sigma}_w = \boldsymbol{\Sigma}_U + n\mathbf{D}_V\boldsymbol{\Sigma}_v\mathbf{D}_V$. By Lemmas B.1(i), B.2(i) and B.3(i), we have

$$
\begin{aligned}
\left\|\boldsymbol{w}^{\intercal}\boldsymbol{w} - \boldsymbol{\Sigma}_w\right\|_{\mathsf{MAX}} \leq\ & \left\|\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{u}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_U\right\|_{\mathsf{MAX}} + 2n^{1/2}\|\mathbf{D}_V\|\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\right\|_{\mathsf{MAX}} \\
& + n\|\mathbf{D}_V\|_{\mathsf{sp}}^2\left\|n^{-1}\sum_{s=1}^{n}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_v\right\|_{\mathsf{MAX}} \\
=\ & O_P\left((\log d/n)^{1/2}\cdot n^{2\bar{\tau}_V^{*+}}\right).
\end{aligned}
$$

For part (ii), by Lemmas B.1(iii), B.2(iii), B.3(ii) and B.3(iii), we have

$$\begin{aligned}
\left\|\boldsymbol{w}^\intercal \hbar \mathbf{D}_H\right\|_{\mathsf{MAX}} &\le \left\|\sum_{s=1}^n \boldsymbol{u}_{s\Delta}\boldsymbol{f}_{s\Delta}^\intercal\right\|_{\mathsf{MAX}} + n\|\mathbf{D}_V\|_{\mathsf{sp}}\left\|n^{-1}\sum_{s=1}^n \boldsymbol{v}_{s\Delta}\boldsymbol{g}_{s\Delta}^\intercal\right\|_{\mathsf{MAX}}\|\mathbf{D}_G\|_{\mathsf{sp}} \\
&\quad + n^{1/2}\left\|n^{-1/2}\sum_{s=1}^n \boldsymbol{u}_{s\Delta}\boldsymbol{g}_{s\Delta}^\intercal\right\|_{\mathsf{MAX}}\|\mathbf{D}_G\|_{\mathsf{sp}} + \|\mathbf{D}_V\|_{\mathsf{sp}}\left\|\sum_{s=1}^n \boldsymbol{v}_{s\Delta}\boldsymbol{f}_{s\Delta}^\intercal\right\|_{\mathsf{MAX}} \\
&= O_P\left((\log d/n)^{1/2}\cdot(1+n^{1+\bar\tau_V+\bar\tau_G}+n^{1/2+\bar\tau_G}+n^{1/2+\bar\tau_V})\right) \\
&= O_P\left((\log d/n)^{1/2}\cdot(1+n^{1/2+\bar\tau_V})(1+n^{1/2+\bar\tau_G})\right) \\
&= O_P\left((\log d/n)^{1/2}\cdot n^{\bar\tau_V^{*+}+\bar\tau_G^{*+}}\right).
\end{aligned}$$

Part (iii) follows from part (ii) as $\left\|\hbar^\intercal \boldsymbol{w}\right\|_1 \le K_H\left\|\hbar^\intercal\boldsymbol{w}\right\|_{\mathsf{MAX}}$. Part (iv) follows from $\|\boldsymbol{w}\|_{\mathsf{sp}} = \left\|\boldsymbol{w}^\intercal\boldsymbol{w}\right\|_{\mathsf{sp}}^{1/2} \le (d\left\|\boldsymbol{w}^\intercal\boldsymbol{w}-\boldsymbol{\Sigma}_w\right\|_{\mathsf{MAX}} + \|\boldsymbol{\Sigma}_w\|_{\mathsf{sp}})^{1/2}$ and $\|\boldsymbol{\Sigma}_w\|_{\mathsf{sp}} \le (\|\boldsymbol{\Sigma}_w\|_1\|\boldsymbol{\Sigma}_w\|_\infty)^{1/2} = \|\boldsymbol{\Sigma}_w\|_1 = O(m_{U,d}+n^{2\tau_V^*}m_{v,d})$. Lastly, we consider part (v). By Lemmas B.1(iv) and B.2(iv), we have

$$\begin{aligned}
\|\boldsymbol{\Lambda}_H^\intercal\boldsymbol{w}^\intercal\|_{\mathsf{sp}} &= \|\boldsymbol{\Lambda}_H^\intercal\boldsymbol{w}^\intercal\boldsymbol{w}\boldsymbol{\Lambda}_H\|_{\mathsf{sp}}^{1/2} \\
&\le (\|\boldsymbol{\Lambda}_H^\intercal(\boldsymbol{w}^\intercal\boldsymbol{w}-\boldsymbol{\Sigma}_w)\boldsymbol{\Lambda}_H\|_{\mathsf{sp}} + \|\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Sigma}_w\boldsymbol{\Lambda}_H\|_1)^{1/2} \\
&\le (K_H\|\boldsymbol{\Lambda}_H^\intercal(\boldsymbol{w}^\intercal\boldsymbol{w}-\boldsymbol{\Sigma}_w)\boldsymbol{\Lambda}_H\|_{\mathsf{MAX}} + \|\boldsymbol{\Lambda}_H^\intercal\|_1\|\boldsymbol{\Sigma}_w\|_1\|\boldsymbol{\Lambda}_H\|_1)^{1/2} \\
&= O_P\left(d^{1/2}(m_{U,d}^{1/2}+n^{\bar\tau_V^*}m_{v,d}^{1/2})(\log d/n)^{1/4}\cdot n^{\bar\tau_V^{*+}} + d^{1/2}(m_{U,d}^{1/2}+n^{\bar\tau_V^*}m_{v,d}^{1/2})\right) \\
&= O_P\left(d^{1/2}(m_{U,d}^{1/2}+n^{\bar\tau_V^*}m_{v,d}^{1/2})(1+n^{\bar\tau_V^{*+}}(\log d/n)^{1/4})\right).
\end{aligned}$$

□

Lemma B.5. *Under Assumptions 1, 3 and 6, we have*
(i) $\left\|n^{-1}\sum_{s=1}^n(n^{-1/2}\boldsymbol{G}_{s\Delta}^c)(n^{-1/2}\boldsymbol{G}_{s\Delta}^{c\intercal})\right\|_{\mathsf{MAX}} = O_P(n^{-1})$;
(ii) $n^{-1}\sum_{s=1}^n \boldsymbol{F}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal} \xrightarrow{d} \int_0^1(\int_0^t\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F - \int_0^1\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F)(\int_0^t\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F - \int_0^1\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F)^\intercal\mathsf{d}t$;
(iii) $\left\|n^{-1}\sum_{s=1}^n(n^{-1/2}\boldsymbol{G}_{s\Delta}^c)\boldsymbol{F}_{s\Delta}^{c\intercal}\right\|_{\mathsf{MAX}} = O_P(n^{-1})$;
where $\boldsymbol{G}_{s\Delta}^c = \boldsymbol{G}_{s\Delta}-\overline{\boldsymbol{G}}$, $\boldsymbol{F}_{s\Delta}^c = \boldsymbol{F}_{s\Delta}-\overline{\boldsymbol{F}}$, $\overline{\boldsymbol{G}} = n^{-1}\sum_{s=1}^n\boldsymbol{G}_{s\Delta}$, *and* $\overline{\boldsymbol{F}} = n^{-1}\sum_{s=1}^n\boldsymbol{F}_{s\Delta}$.

Proof. Parts (i) and (ii) are trivial. For part (iii), by Lemma B.3(v), we have

$$\left\|n^{-3/2}\sum_{s=1}^n\boldsymbol{G}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal}\right\|_{\mathsf{MAX}} = n^{-1/2}\left\|n^{-1}\sum_{s=1}^n\boldsymbol{G}_{s\Delta}\boldsymbol{F}_{s\Delta}^\intercal - \overline{\boldsymbol{G}}\,\overline{\boldsymbol{F}}^\intercal\right\|_{\mathsf{MAX}} = O_P(n^{-1}).$$

□

Lemma B.6. *Under Assumptions 1–8, we have*
(i) $\|n^{-1}\mathscr{H}^{*c\intercal}\mathscr{H}^{*c}\|_{\mathsf{sp}} = O_P(1)$*;*
(ii) $\left\|\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal - \mathscr{H}^c\right\|_{\mathsf{MAX}} \leq \left\|\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal - \mathscr{H}^c\right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right)$*;*
(iii) $n^{-1}\|\boldsymbol{\Xi}\widehat{\mathscr{H}^{*c}}^\intercal\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal - \mathscr{H}^{*c\intercal}\mathscr{H}^{*c}\| = O_P\left(n^{-1/2-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right)$*;*
where $\boldsymbol{\Xi}$ *is defined in Lemma 3.1.*

Proof. (i) By Lemma B.4, the dominate term is $n^{-1}\sum_{s=1}^n \boldsymbol{F}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal}$, which is of order $O_P(1)$.(ii) By Theorem 3.1, we have

$$\left\|\widehat{\mathscr{H}^*}\boldsymbol{\Xi}^\intercal - (\mathscr{H} - \mathbf{1}_n\boldsymbol{h}_0^\intercal)\right\|_{\mathsf{MAX}} \leq \left\|\widehat{\mathscr{H}^*}\boldsymbol{\Xi}^\intercal - (\mathscr{H} - \mathbf{1}_n\boldsymbol{h}_0^\intercal)\right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right).$$

Since $\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal = (\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^\intercal/n)\widehat{\mathscr{H}}\boldsymbol{\Xi}^\intercal$, $\mathscr{H}^c\boldsymbol{\Xi}^\intercal = (\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^\intercal/n)(\mathscr{H} - \mathbf{1}_n\boldsymbol{h}_0^\intercal)$ and $\|\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^\intercal/n\|_{\mathsf{sp}} = 1$, we then have $\left\|\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal - \mathscr{H}^c\right\|_{\mathsf{sp}} = O_P\left(n^{-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right)$.
(iii) The result follows by noticing that

$$\begin{aligned}
&\|\boldsymbol{\Xi}\widehat{\mathscr{H}^{*c}}^\intercal\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal - \mathscr{H}^{*c\intercal}\mathscr{H}^{*c}\| \\
\leq\ & \left\|\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal - \mathscr{H}^c\right\|_{\mathsf{sp}}\|\mathscr{H}^c\|_{\mathsf{sp}} + \left\|\widehat{\mathscr{H}^{*c}}\boldsymbol{\Xi}^\intercal - \mathscr{H}^c\right\|_{\mathsf{sp}}^2 \\
=\ & O_P\left(n^{1/2-2\underline{\tau}_G^{*-}} \cdot a_{nd}\right).
\end{aligned}$$

□

Figure 1: Estimated number of factors for each trading day from 11th January, 2021 to 14th May, 2021. The y-axis represents the number of factors and the x-axis represents the dates (given in the format mmdd). The y-coordinate of the top of each grey bar gives the estimated total number of factors, $\widehat{K}_H$, from IC_1. The length of each grey bar represents the difference between $\widehat{K}_H$ and $\widehat{K}_F$, which is obtained from the PANIC test using 1% significance level. The length of each red bar represents the difference between $\widehat{K}_F$'s obtained from the PANIC tests using 1% and 5% significance levels. The length of each blue bar represents the difference between $\widehat{K}_F$'s obtained from the PANIC tests using 5% and 10% significance levels. The y-coordinate of the bottom of each blue bar gives the value of $\widehat{K}_F$ obtained from a 10% PANIC test.

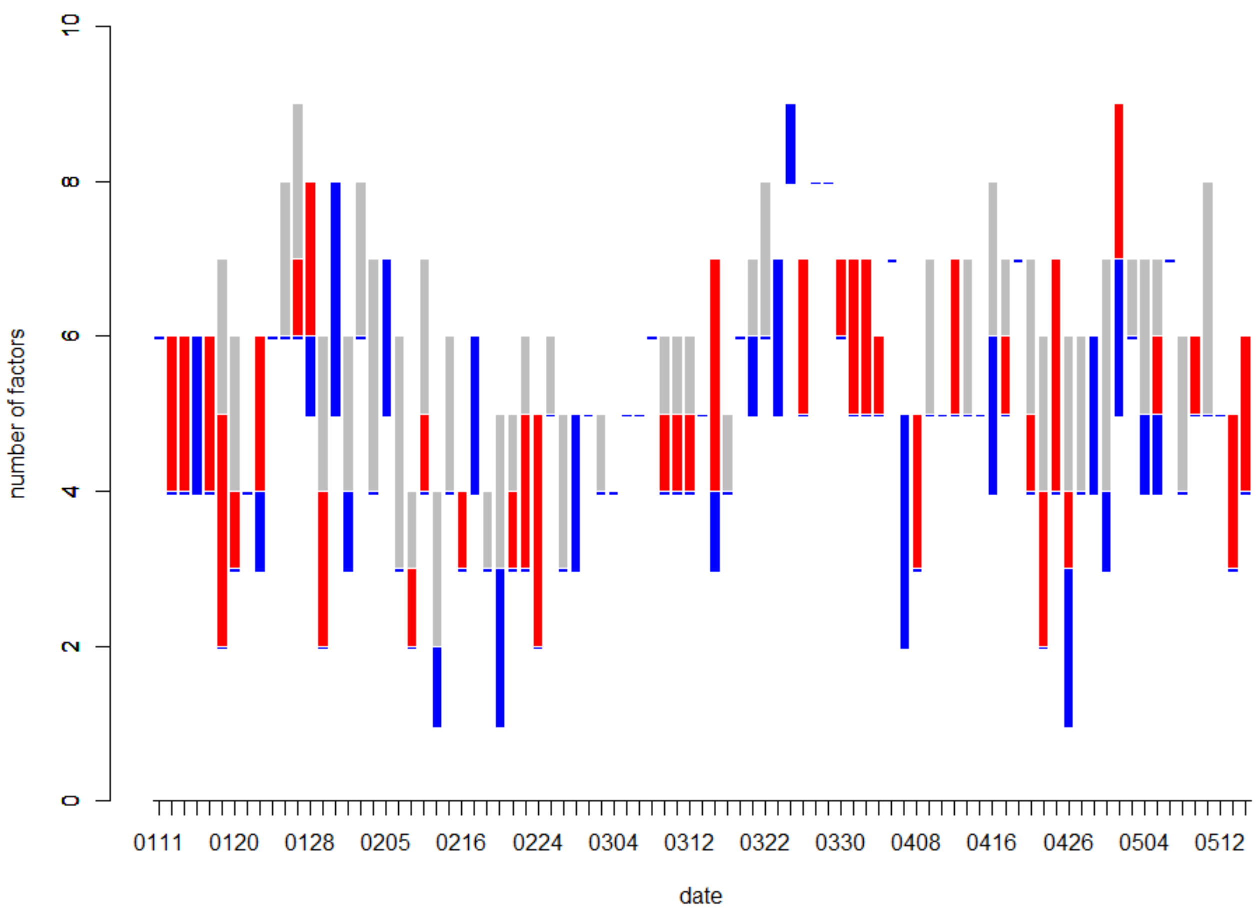

Figure 2: Plot of the six cumulative factors estimated from the DPCA*, i.e., plot of the six components of $\left(\widehat{\boldsymbol{\beta}}_{\perp}, \widehat{\boldsymbol{\beta}}\right)^{\intercal} \widehat{H}^{*}_{s\Delta}$, $s = 1, \ldots, 78$, for data on 26th April 2021.

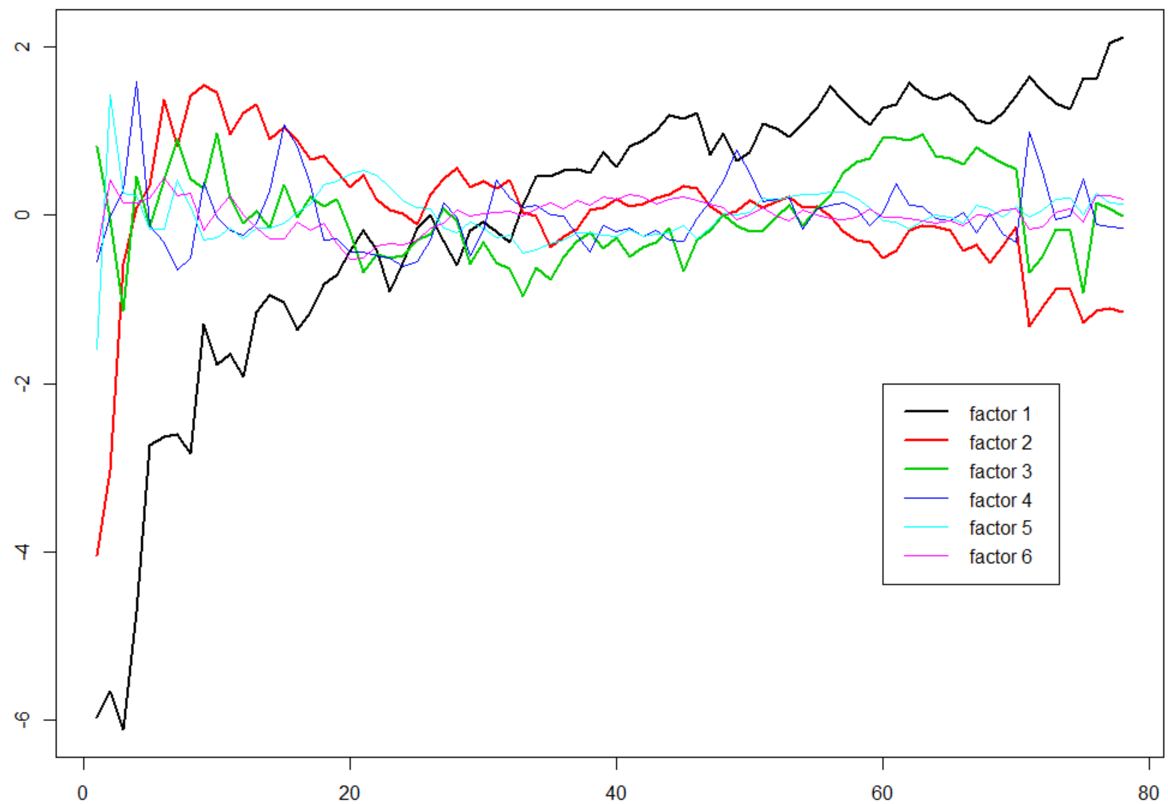

Figure 3: Decomposition of aggregated returns into the common components of efficient prices (CC.EP), the common components of microstructure noise (CC.MN), and the idiosyncratic errors (Residual) for the stocks A, AAL, AAP, AAPL, and ABBV when $\widehat{K}_F = 4$ and $\widehat{K}_G = 2$. Each row gives the decomposition for each stock, with the first diagram giving the aggregated returns, followed by CC.EP, CC.MN, and Residuals.

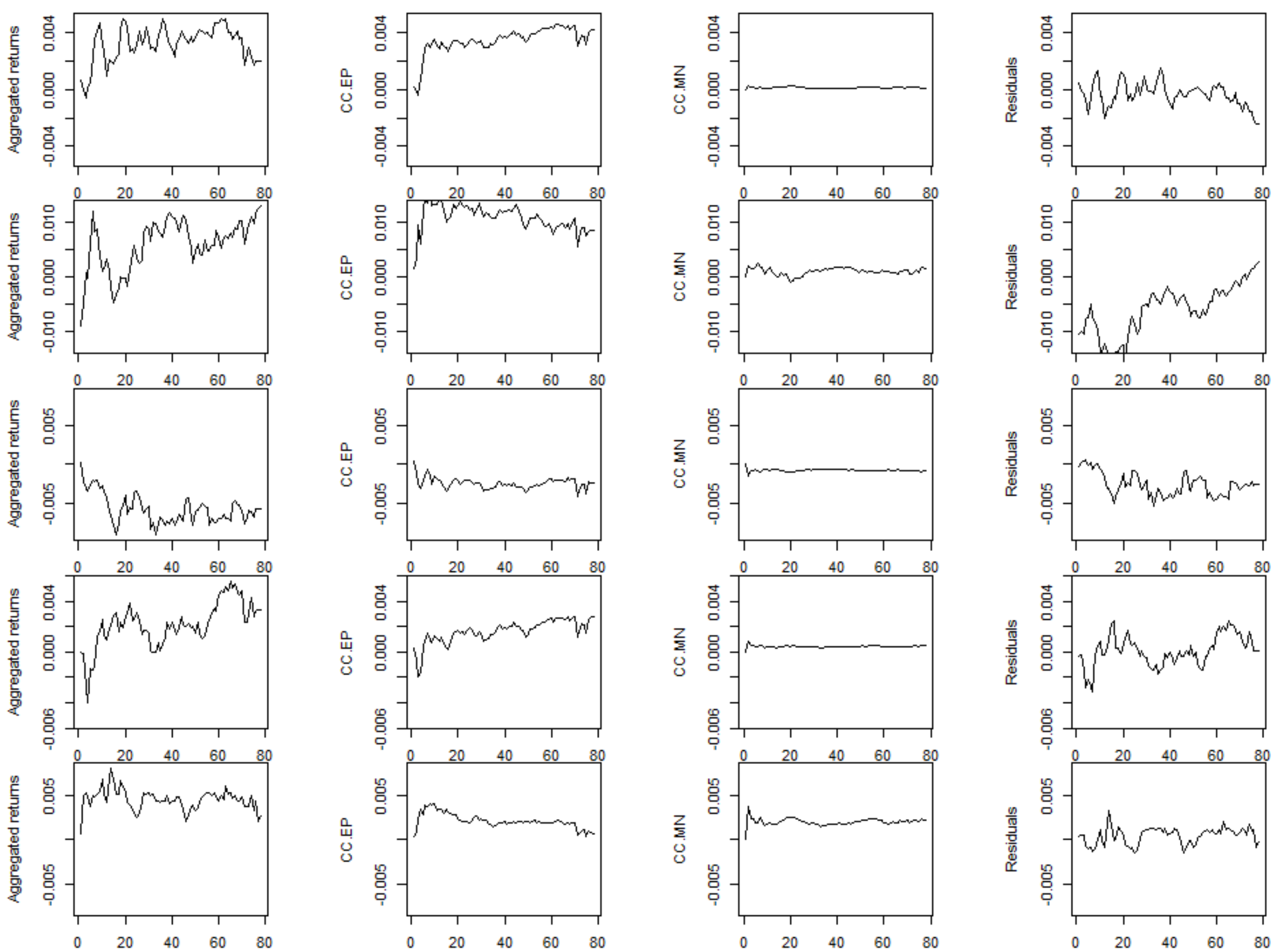

Figure 4: Decomposition of aggregated returns into the common components of efficient prices (CC.EP), the common components of microstructure noise (CC.MN), and the idiosyncratic errors (Residual) for the stocks A, AAL, AAP, AAPL, and ABBV when $\widehat{K}_F = 3$ and $\widehat{K}_G = 3$. Each row gives the decomposition for each stock, with the first diagram giving the aggregated returns, followed by CC.EP, CC.MN, and Residuals.